INFORMATION PROCESSING MODELS AND COMPUTER AIDS FOR HUMAN PERFORMANCE

SECOND-LANGUAGE LEARNING

BOLT BERANEK AND NEWMAN, INCORPORATED

PREPARED FOR

AIR FORCE OFFICE OF SCIENTIFIC RESEARCH

ADVANCED RESEARCH PROJECTS AGENCY

30 JUNE 1974

UNCLASSIFIED

Security Classification

AD787876

DOCUMENT CONTROL DATA - R & D

(Security classification of title, body of abstract and indexing annotation must be entered when the overall report is classified)

1. ORIGINATING ACTIVITY (Corporate author)

BOLT BERANEK AND NEWMAN INC.
50 Moulton Street
Cambridge, Mass. 02138

2a. REPORT SECURITY CLASSIFICATION

unclassified

2b. GROUP

3. REPORT TITLE

INFORMATION PROCESSING MODELS AND COMPUTER AIDS FOR HUMAN PERFORMANCE: SECOND-LANGUAGE LEARNING

4. DESCRIPTIVE NOTES (Type of report and inclusive dates)

Final Report, covering activities to 30 June 1974.

5. AUTHOR(S) (First name, middle initial, last name)

Daniel N. Kalikow

6. REPORT DATE	7a. TOTAL NO. OF PAGES	7b. NO. OF REFS
30 June 1974	~~150~~ 152	8

8a. CONTRACT OR GRANT NO.

F44620-71-C-0065 (ARPA)

b. PROJECT NO.

AO 1993

c. 61101E

d. 681373

9a. ORIGINATOR'S REPORT NUMBER(S)

9b. OTHER REPORT NO(S) (Any other numbers that may be assigned this report)

AFOSR - TR - 74 - 1730

10. DISTRIBUTION STATEMENT

Approved for public release; distribution unlimited.

11. SUPPLEMENTARY NOTES

12. SPONSORING MILITARY ACTIVITY

Air Force Office of Scientific
1400 Wilson Boulevard (NL)
Arlington, Virginia 22209

13. ABSTRACT

This report describes the second field evaluation experiment on the Mark II model of the Automated Pronunciation Instructor (API) system. Two matched groups of students were studied. All were native speakers of Spanish, and all were enrolled in the Intensive English Program at the University of Miami. One group was tested and trained with the API system; the other was simply tested within the same time frame. Students exposed to the system showed no greater improvement in the ability to discriminate the sounds of the target language than did the control students over the same period of time. The experimental students did show significantly more improvement than their control counterparts in a speech production test. However, the size of the treatment effect was not sufficient to demonstrate the cost-effectiveness of the API system in its present form.

152

DD FORM 1 NOV 65 1473

UNCLASSIFIED

UNCLASSIFIED

Security Classification

14 KEY WORDS	LINK A		LINK B		LINK C	
	ROLE	WT	ROLE	WT	ROLE	WT
speech analysis						
language learning						
computer-assisted instruction						
man-computer interaction						

Report No. 2841 Bolt Beranek and Newman Inc.

INFORMATION PROCESSING MODELS AND COMPUTER AIDS FOR HUMAN PERFORMANCE

FINAL REPORT

SECOND LANGUAGE LEARNING

30 June 1974

by

Daniel N. Kalikow

ARPA Order No. 1993

Sponsored by the
Advanced Research Projects Agency
Department of Defense
under
Air Force Office of Scientific Research
Contract F44620-71-C-0065

Prepared for
Air Force Office of Scientific Research
1400 Wilson Boulevard
Arlington, Virginia 22209

FINAL REPORT

30 June 1974

ARPA Order No. 1993

Program Code No. 3D20

Contractor: Bolt Beranek and Newman Inc.

Effective Date of Contract: 1 January 1971

Contract Expiration Date: 30 June 1974

Amount of Contract: $689,260

Contract No. F44620-71-C-0065

Principal Investigators: John A. Swets

Daniel N. Kalikow

Telephone No. 617- 491-1850

Title: INFORMATION PROCESSING MODELS AND COMPUTER AIDS FOR HUMAN PERFORMANCE

TABLE OF CONTENTS

TABLE OF CONTENTS (cont)

SUMMARY

1. Technical Problem

The task is to carry out the final development of a computer-based system for automated instruction of the new speech sounds of second languages, and to field-test this system for two language pairs: English speakers learning Mandarin Chinese, and Spanish speakers learning English.

2. General Methodology

Laboratory experiments and field evaluations.

3. Technical Results

This report describes the second field evaluation experiment on the Mark II model of the Automated Pronunciation Instructor (API) system. Two matched groups of students were studied. All were native speakers of Spanish, and all were enrolled in the Intensive English Program at the University of Miami. One group was tested and trained with the API system; the other was simply tested within the same time frame. Students exposed to the system showed no greater improvement in the ability to discriminate the sounds of the target language than did the control students over the same period of time. The experimental students did show significantly more improvement than their control counterparts in a speech production test. However, the size of the treatment effect was not sufficient to demonstrate the cost-effectiveness of the API system in its present form.

4. Department of Defense Implications

Language schools of the Department of Defense give instruction in approximately 65 languages to over 200,000 students each year. The systems under development are designed to facilitate this instructional process.

ACKNOWLEDGMENTS

The author gratefully acknowledges the contributions of the other members of the project team: Juan L. Anguita, Douglas W. Dodds, Kenneth N. Stevens, Ann M. Rollins, and John A. Swets. The advice of Raymond S. Nickerson is greatly appreciated.

Particular thanks are expressed to the faculty, staff, and students of the University of Miami Intensive English Program, notably Dean M. Robert Allen and Mr. John H. Rogers. Their gracious hospitality and unfailing interest made our work very pleasant.

The advice of Dr. Austin Kibler of ARPA and of Dr. Charles Hutchinson of AFOSR is also acknowledged.

PREFACE

The present contract is a partial continuation of a research program begun in 1966 under ARPA sponsorship. Of the four tasks at one time funded under AFOSR contract F44620-67-C0033, the present task remains active under Contract F44620-71-C0065. The present Technical Report covers activities undertaken during the 1973-74 academic year, during which the Automated Pronunciation Instructor (API) system was field-tested at the University of Miami's Intensive English Program. The source language of selected students was Spanish.

The Spanish-English language pair field test is the second and last in the program funded under the present contract. Previous technical reports have described the development and construction of system hardware and software, and have presented the results of a smaller-scale field test undertaken in Cambridge for the English-Mandarin Chinese language pair. The present report has three overall aims.

1. To place all previous work in an integrated presentation, allowing references to be made within the bounds of one report to all of the work performed under this contract; to develop the rationale for the last field-test.

2. To present and discuss the final field-testing trials.

3. To discuss the body of knowledge generated by this work, and to summarize the results obtained.

1. BACKGROUND AND INTRODUCTION

1.1 Previous Research

Work on the development of the Automated Pronunciation Instructor (API) system has been done in two major stages over the past few years. The first stage culminated in the development and testing of the Mark I version of the API system for the Spanish-English language pair; the second stage, whose conclusion is the main subject of the present report, consisted in the redesign of the API and its testing in two language pairs.

1.1.1 Mark I API, Spanish-English Experiment

Kalikow and Swets (1972) described the hardware and software of this system, and summarized the results of experiments undertaken in Cambridge with local Spanish-speaking housewives learning English through exposure to their English-speaking environment. Subjects designated "experimental" were exposed to both audio and visual feedback from the API system; "control" subjects worked with the same material and had the benefit of the audio feedback, but not the visual - the central innovation of the system. All subjects were paid for their participation, and were exposed to the system for 90 minutes/week for 8 weeks. They were not concurrently studying English in any formal academic sense.

The Mark I API provided audiovisual feedback for the following accent problems in the Spanish-English language pair:

a. Vowels in monosyllabic words, using a schematic "tongue-position" cross-section display of the mouth, with time as the parameter of the display.

b. Reduced vowels in multisyllabic words, using the above, "tongue-position" display in conjunction with a syllable-finder.

c. Aspirate/Unaspirate initial stops, using a display showing the presence and intensity of aspiration noise, and the presence and duration of the time invervening between word and voicing onsets.

The latter display was unique in that it made time an explicit part of the display, and in that it also provided evaluational criteria for student use. Students reported this type of display as easy to understand.

System effectiveness was assessed by having both groups of subjects read a list of words similar to those trained, both before and after training, and after a retention (no-training) interval. Selected utterances were removed from the testing tapes and pairwise compared (within subjects and words, and across testing times). Strong training effects were observed for both experimental and control subjects, leading to the tentative conclusion that the control treatment was too convervatively designed. The presence of the training effects, the comments of the subjects, and the differential effectiveness of the various displays both encouraged and directed subsequent work.

1.1.2 Mark II API, English-Mandarin Chinese Experiment

The system was designed around a PDP8-E computer, and reprogrammed in such a way as to greatly increase its capabilities for speech analysis and display. A full description of hardware and basic system software was given in Kalikow (1972). The major differences obtaining between Marks I and II are the following:

a. The invariable use of time as an explicit dimension of the display; i.e., speech parameters extracted are displayed directly against an abscissa time-base.

b. The provision for direct analysis of prerecorded teacher speech. Whereas previous displays had incorporated "targets," these were abandoned in the newer version for their lack of explicitness and occasional gross inaccuracy. The teacher's speech is analyzed and displayed in the same manner as is the student's.

c. Increased capacities for speech storage -- both analog and digital, and generally enhanced interactive capabilities.

d. Capability for direct pitch extraction from the speakers.

e. Use of minimal pairs and phrases where appropriate, rather than only single words.

The planned course of the development and evaluation of the Mark II API originally envisioned three phases. First, system construction and checkout, with concurrent software development. Second, field trials of the system in the Spanish-English language pair. Finally, field trials in the English-Mandarin Chinese language pair. The field trials were to be carried out within the instructional facilities, and following the curriculum guidelines, of the Defense Language Institute. The subsequent unavailability of this theater of operations necessitated reordering the schedule.

The first test of the new system was carried out using English-speaking students of Mandarin Chinese from two neighboring universities (Kalikow and Rollins (1973). Two groups of seven students, matched according to their pronunciation abilities, were formed and pretested with a list of utterances similar to those to be trained. Experimental students were then allowed to work with the API, and control students were simply retested within the same time frame as the experimentals, while

continuing their parent course in Chinese. Both groups re-read the test list following conclusion of 8 training sessions for the experimentals, and following a four-week retention interval.

The training curriculum for this experiment emphasized two major problems facing the student of this language pair.

a. Tone production. Utilizing the pitch detector and the time display of fundamental frequency, students compared their tone contours with those of a teacher in a graded series of utterances and display options. The simplest utterances were single tones; the most complex utterances were two-syllabic tone groupings where "tone sandhi" exerts its complicating effects on the fundamental frequency contour.

b. Aspirate and unaspirate voiceless initial stops. Four difficult contrasts were chosen, each involving articulator positions and/or temporal energy distributions that are not present in English. The display was a composite of the pitch trace to give feedback on the presence and contour of voicing, added to which was a loudness trace, to give feedback on the presence, magnitude, and onset time of voiceless sounds.

As in all API research, the system did not specifically evaluate student utterances, but provided visual pattern feedback tied to relevant speech parameters, in conjunction with audio (analog tape loop) feedback of student and teacher speech. The students' pattern-recognition capabilities were enlisted in self-evaluation and speech modification.

Utterances from the list recorded by the students on the three test days were presented in randomized orders to a group of Mandarin-native teachers. They rated the speech on a five-point scale of accentedness. An improvement in speech was defined as a positive-going difference in one judge's ratings for a given utterance by the same subject over time. Overall, and in several word groupings, the experimental subjects demonstrated significantly superior rates of improvement when compared with the control group.

Despite the restrictions forced by the necessity of performing the research in our laboratory rather than in closer proximity both to students and the parent course, measurable improvements in Chinese pronunciation were obtained. It remained to be proven whether such positive indications could be said to be significant in the broader sense of genralized improvements in target-language skills. Such questions can only be effectively addressed if the research is itself carried out on a broader scale. Both the Chinese and the initial Spanish-English experiments had used students visiting at most twice a week, with at best little interfacing to a parent course in the target language. The problem of specifying the control treatment was not solved in the same manner in these two experiments, and might have erred first in too conservative a direction, and in too liberal a direction for the Chinese test.

1.2 The University of Miami Intensive English Programs

The site chosen for the Spanish-English field test provided a good combination of factors favoring the research; adequate student sample, faculty receptivity, curricular adaptability, administrative cooperation, and logistics. The Intensive English Program is located on the Coral Gables campus of the

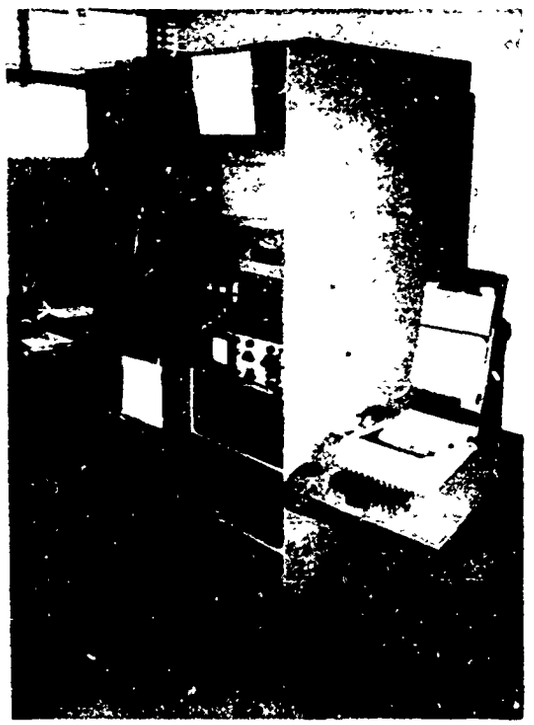
University of Miami. Approximately 100 foreign students per semester enroll, with the eventual aim of studying at an American university. About 30% of the students are of Spanish-speaking background. Appendix 1 gives some details of the organization, administration, and regulation of this curriculum.

The class assignments are made on the basis of student performance on an initial test emphasizing vocabulary and grammar, or on the basis of past performance for returning students. Students are grouped into three general levels of course difficulty. A balanced curriculum is taught over the 15-week semester, including grammar, conversation, pronunciation (classroom drills and discussion), reading, writing, conversation, and language laboratory. Overall, the program places more emphasis on pronunciation skills than most second-language curricula, with some interesting instructional innovations not found elsewhere.

One of the major attractive features of the IEP was its pre-existing use of the language laboratory. While the facility itself does not have provision for student-controlled recording and playback, drill curricula interfacing rather well with the API approach have been generated and used successfully by the Director of the IEP language laboratory (Balian, 1972). Particularly in the area of vowel pronunciation, the minimal-pair approach is heavily used, with attention paid to the orthographic confusions that often plague the foreign student of English. In summary: the instructional environment was hospitable; the students were being exposed to a fair approximation of state-of-the-art pronunciation training; therefore, the Intensive English Program provided a fair background against which potential API effects might be evaluated.

An agreement was reached with the administration and staff of the University of Miami and of the Intensive English Program regarding the use of the API with their students. It was agreed, among other things, that students were to utilize the API in addition to their normal class work, and that the API was not to substitute for any portion thereof. Experimental students would have to be trained in their free time; control students could be made available on the same basis, but the question of what treatment would be appropriate for them was left open.

The API system was installed in the building housing the Intensive English Program. See Figure 1. This arrangement placed the teaching machine in the middle of the classroom layout, facilitating the use of the API during free periods. The work was planned to encompass two successive semesters, each semester's students being selected, trained, and tested in as similar a manner as possible. This was the major means available to increase the number of students using the API, because there was an outside limit on student participation, placed by the number of hours available at the intersection of three sets of time variables: normal business hours at the class building, free times of students in the various class sections, and the other duties of the field engineer who sets up the system for each student's training period.



A. Computer System

B. Student seated at work table within sound-treated enclosure.

Figure 1. Illustrating the API system as set up within the University of Miami Intensive English Program Building.

2. METHOD

2.1 Experimental Design Summary

Because the experiment to be described encompasses a large amount of time and has several interlocking procedures, it is useful to sketch the overall approach to aid the reader in achieving an integrated view of the process. The dependent variable being measured here is "accentedness" of speech in the target language, or its presumed correlates. The independent variable is exposure to the API system. The means of controlling for exposure variables, and of measuring the accent variables, are at the heart of the experimental design; and because this work was undertaken in a real-world environment, certain suboptimal arrangements were sometimes necessary.

The first requisite for the experiment is an adequate number of students having the appropriate language background. From this population, two matched groups are formed. The matching is to be done through the use of measuring instruments (tests) that are independent of the means to be used in training the eventual experimental group. These pre-training measures are to be used, therefore, as a means for equating the pronunciation-related skills of the two groups. As a part of the pretesting of the groups, a recording is made of each student as he reads materials similar to those upon which the experimental students will be trained.

The treatment phase is entered next. Experimental students are trained regularly on the API system, using a curriculum of graded difficulty in terms both of accent-reduction problems and of display subtlety. Control students are given printed copies of the training lists used by the experimentals, and are encouraged to work with those materials.

Immediately following the conclusion of the training of the experimentals, both groups are "post-tested" using the same test battery as used at the outset. No contact is then made with the students for four weeks, as they continue their work in the parent course. Immediately prior to the conclusion of the parent course, a "retention-test" of the same type as the previous two is given again to all students. The problem of re-exposure of the same students to the same test instruments is lessened in severity in this design. It might be said that the students' scores might be artifactually increased through simple experience with the measures themselves, but this has little impact if the data are analyzed differentially. In the absence of "alternate forms" of the accent-related tests, the data from repeated exposures to the same test can still be useful if group performance differences can be traced to training differences, over and above those differences presumably caused primarily by simple exposure to the test itself. Presumably, the control treatment specified above is a reasonable approximation to that state of affairs.

The data generated are analyzed by the appropriate objective and subjective means, and differential statistics are utilized to evaluate the reliability of any differences observed that are traceable to the experimental treatment. Specific details on all phases of activity introduced above will be provided next.

2.2 Student Selection

All new incoming students of the Intensive English Program are given a placement test of the "fill-in-the blanks" type, emphasizing vocabulary, grammar, and common English usage. Returning students are usually simply assigned to the appropriate class section. Different sections have different daily schedules, each involving

five class periods, and it was therefore necessary to select students from the various sections so as to allow maximum utilization of the API on a daily basis. Data from the Intensive English Program placement test were, therefore, used as part of the selection process. Returning students of Spanish background were also given the standard placement test in this case, to complete the data base.

The most important datum on each potential student was the score on a discrimination test designed to point up the pronunciation difficulties peculiar to the Spanish-English language pair. Lado's Test of Aural Perception in English for Latin-American Students (1957) was used to provide information presumably closely related to pronunciation skills. It is important to be able to demonstrate that, for an experiment designed to emphasize post-training differences between an experimental and control group, the groups were matched before the application of the experimental treatment. The assessment of actual pronunciation skills is quite cumbersome, as will be seen in the presentation of the analysis of the recorded utterances of the students. There is not the time at the start of the Intensive English course to perform the laborious process of speech analysis for the purposes of group formation. The Lado test produces data that are quickly accessible, and useful both for group balancing and for post-training evaluation as well.

All incoming students were administered the Lado test. Scores were given to all interested students upon request. In the two terms, a total of 38 Latin-American students was studied through to the conclusion of the

treatments. These students were assigned to the experimental or control groups in pairs. Figure 2 illustrates the scattergram of assignment data, and Table 1 identifies the subjects and demonstrates the similarity in their English abilities in these two tests administered at the outset of training. Because the methods and curriculum were so similar in both terms, it was possible to combine data from the two terms, and even to select members of a given "matched pair" from students in two different terms. In each term, all available incoming students of Spanish background were utilized, with the maximum number of them to receive the experimental treatment. Similar scattergram selection procedures were employed to balance these groups as closely as possible, but the data for individual terms are not presented here due to the complicating effects of student dropouts from both the experimental and control groups. In the second term, several extra control students were initiated with the express purpose of pairing them with first-term experimentals whose first-term controls had withdrawn. In both terms, assignment to a treatment group was determined both by the attempt to pair students whose locations in Fig. 2 were maximally close, and by the scheduling exigencies imposed by the parent course. It was often impossible to resolve such conflicts, with the result that the latter criterion governed; it had been agreed at the outset that service as an experimental student was not to interfere with the normal course work. The final arrangement of Fig. 2 reflects these original decisions in most cases, but when the members of a given pair are drawn from two terms, this was a post-hoc decision. Because it is inappropriate to use matched-pair statistics on such data due to this mixture of assignment procedures, none is attempted below. The purpose of scattergram usage was to derive group assignments in a realistic and efficient manner, in such a way as to permit the statement that the two groups are insignificantly different in their

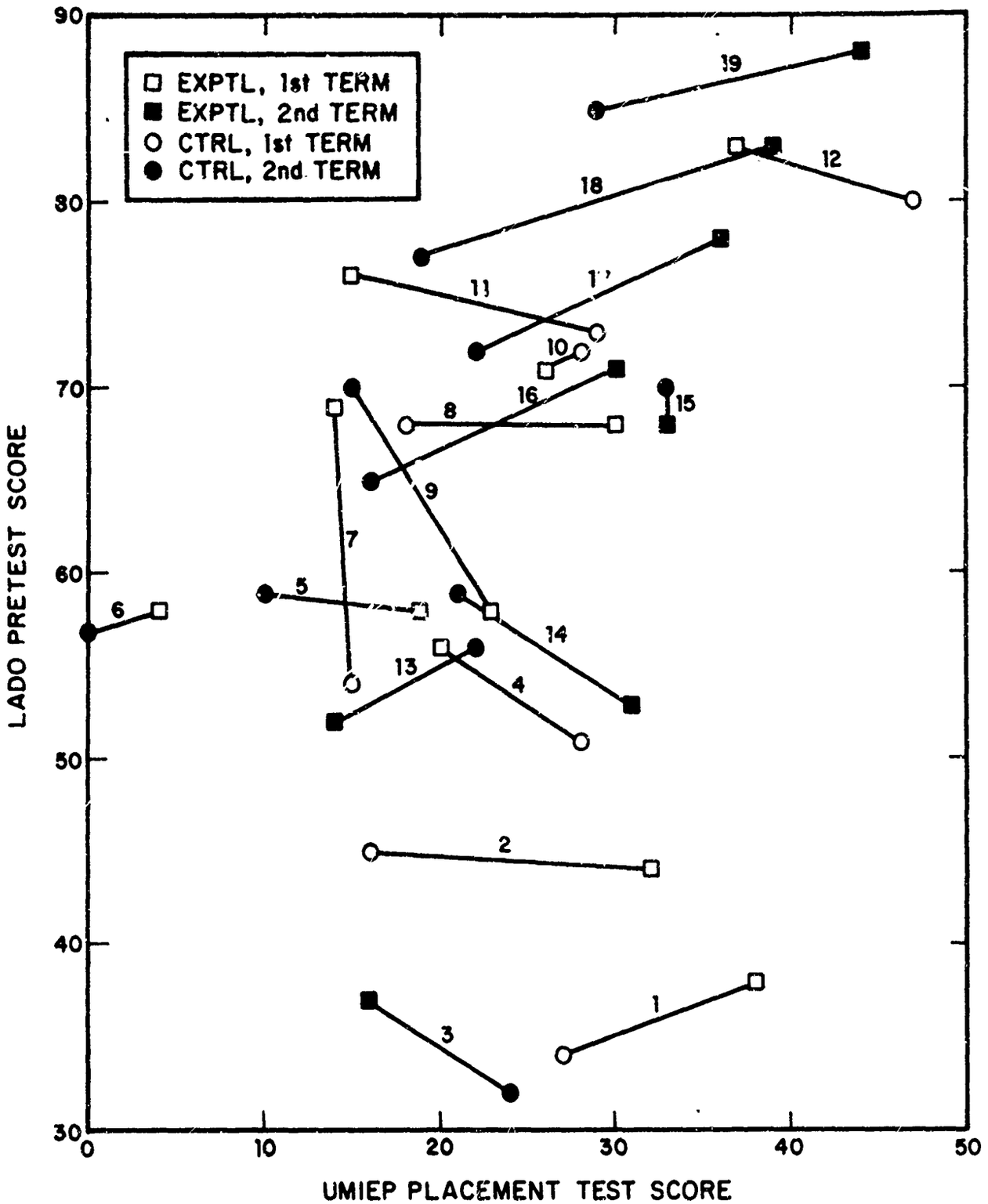


Figure 2. Scattergram of Matching Data.

TABLE 1

Matching Data

EXPERIMENTAL GROUP	UNIEP Placement	LADO PRE	CONTROL GROUP	UNIEP Placement	LADO PRE
E1. Pomponio, I.	38	38	C1. Lopez, L.	27	34
E2. Gongora, A.	32	44	C2. Vegas, M.	16	45
E3. Mora, M.	16	37	C3. Luz, N.	24	32
E4. Peralta, V.	20	56	C4. Zuluaga, A.	28	51
E5. Bild, A.	21	58	C5. Szwarl, B.	10	59
E6. Cespedes, M.	4	58	C6. Vivas, S.	0	57
E7. Morelli, B.	14	69	C7. Franco, G.	15	54
E8. Demoya, T.	30	68	C8. Bermudez, I.	18	68
E9. Gomez, J.	23	58	C9. Orellana, R.	15	70
E10. Vegas, C.	26	71	C10. Priewer, C.	28	72
E11. Hernandez, C.	15	76	C11. Castellanos, M.	29	73
E12. Farberoff, J.	37	83	C12. Alalu, F.	47	80
E13. Compan, A.	14	52	C13. Kaplan, D.	22	56
E14. Acosta, M.	31	53	C14. Ruiz, C.	21	59
E15. Bermudez, I	33	68	C15. Gonzalez, E.	33	70
E16. Abumohor, A.	30	71	C16. Lechuga, C.	16	65
E17. Puga, M.	36	78	C17. Restrepo, M.	22	72
E18. Fraynd, P.	39	83	C18. Cocchis, A.	19	77
E19. Castellanos, M.	44	88	C19. Steinworth, W.	29	85
Means:	26.5	63.6		22.0	62.0

Underlined Subject code numbers indicate that the individual was run in the Spring semester.

pre-training English abilities. The means and raw data dispersions shown in Table 1 demonstrate the reasonableness of these assignments.

2.3 Orientation and Pretesting

Following selection, all students were gathered for an orientation meeting. They were told of the purpose of the experiment, of the operations of the API, and of their roles in the coming project. The uniform reaction of the control students was disappointment, and eagerness to participate in a more active role. Experimental and control students were each given a complete listing of the curriculum lists (described below) to be used on the API, and the controls were particularly exhorted to study them carefully. This was to be the control treatment. It had proven infeasible to arrange any realistic pronunciation training for the control group. Exposure to the API without visual feedback was impossible; its time was already fully committed to instruction of experimental students. So, too, was the time of the field engineer. Organized pronunciation drills using the API curriculum were theoretically possible, but the distribution of free times across the controls was not appropriate, and the issue of exactly what would be an appropriate control treatment could not be adequately resolved. In the end, we depended on the natural curiosity of the students, and on a friendly rivalry between the two groups. The similarity of the training curriculum to the normal language laboratory fare facilitated the control student's study of the API materials.

Speech samples were collected from all students prior to the start of training for the experimental group. The script for this procedure is included as Appendix 2 of this report. The recording was made in the sound-treated student booth. Materials read

covered the subject matter to be trained, but were different from the training materials themselves. This was to reduce the later test-day advantage to the experimental students that might obtain simply because of increased exposure to the training materials. The testing (i.e., voice-sampling) situation was made as dissimilar as possible to the normal training configuration of the API, also in an attempt to reduce the later familiarity advantage of the experimental students. The field engineer sat in the booth with the student, prompting when necessary with instructions to re-read items produced with improper inter-item intervals or with art factual noises near or within the utterances. He did not comment on improper pronunciations.

Since English orthography is a potent source of confusion for the new student, every effort was made to ease his task. For this purpose, the vowel-cuing system developed by Balian was included in the script, and increasingly utilized as the students gained experience in the course. Students unfamiliar with the numbers and symbols written near the test items in the script were told to ignore them. Naturally, they were to take more advantage of those cues upon later retesting on the same script; but since experience was being controlled, the bias would tend to be uniform across groups.

The four pages of voice-sample material were written with the following aims. Page 1 was simply for acclimatization of the student; those data were not to be analyzed. Page 2 contained illustrative utterances for evaluation of intonation contours, rhythm, and general naturalness of the speech. Page 3 was included in the hopes of testing production of certain vowel and consonant contrasts in running speech. The same minimal pair items to be tested in relative isolation on page 4 were

incorporated within a set of rather contrived utterances. Finally, page 4 contained minimal pairs covering the vowels and consonants to be trained. These were read as quartets, again to minimize similarities between API training utterance format and the testing situation.

2.4 Training Procedures

2.4.1 Curriculum

The materials upon which the students were trained are shown in Appendix 3. An identical booklet was given to each experimental and control student. Each of the lists was recorded on API teacher cartridges by two speakers of the General American dialect (DNK and DWD). Separate versions of the teaching software were available on the digital storage device within the API, and so in order to train a given student with a particular word-list, the field engineer had only to mount the appropriate teacher cartridge and load the corresponding software into the computer itself. From that point, after some initial equipment settings, the student worked independently of the monitor.

The accent problems addressed by each of these lists, and the display algorithms used, are contained therein. Some additional comments are needed in explanation of the format of individual lists and of their interrelationships. The intonation lists were derived largely from University of Miami language laboratory scripts and other curriculum materials. The balance of the API curriculum lists is in the minimal pair format, and addresses specific vowel and consonant contrasts as indicated. One member of each of the vowel pairs is always one of the English vowels maximally similar to one of the "home vowels" of Spanish; the other member of the pair is a new vowel presumably causing trouble

for large numbers of Spanish-speaking students. Similarly, the consonant pairs each contain a difficult member contrasted with an easily produced consonant. Lists whose codes terminate with "A" are arranged such that the easier item is always first in the trained utterance. Lists whose codes terminate in "B" have been arranged such that the order of difficulty alternates through the list. These "scrambled" versions were introduced later in the training regimen, as outlined below.

2.4.2 Student Orientation

The students' first training session was carried out with the field engineer in attendance. He demonstrated the proper modes of system operation, described the important aspects of the display, and in general assured himself that the student was in full command of all system features. His Spanish-language capabilities were invaluable at this time, since there was no language barrier interposed for this fundamentally important session. The list coded INT 1 DWD was used.

At a later point in training, when deemed ready by the monitor, each student was given a copy of the booklet "Hints for the Student," attached to the present report as Appendix 4. This gives complete instructions on display interpretation, as well as many illustrative examples of display output. It would be repetitious to recapitulate those instructions here. Students required very few additional instructions by the monitor before becoming completely conversant with system use. At no time did the monitor explicitly evaluate student or system performance.

2.4.3 Instructional calendar and display modifications through training.

During the two semesters, a total of 19 experimental students was trained: twelve in the first semester, and seven in the second. Each student worked with the system at a fixed time each day, five days per week, 45 minutes per day. The schedule of the Fall semester permitted 42 training sessions (not counting the orientation session), while in the Spring semester, only 37 sessions could be accommodated before a long vacation intervened. The aim was, of course, to train the subjects in a continuous manner and post-test them immediately; the parent-course instructional content during the retention interval was irrelevant, and the vacation provided a convenient breakpoint.

Table 2 illustrates the manner of passage through the curriculum. It was done in three passes: initial, scrambled recapitulation, and review. For all three passes, the MATCH function operated in a consistent manner within each of the three types of curriculum material. For the Intonation lists, where complete phrases and sentences were trained, vertical phrase matching was provided. For the Vowel lists, sliding (i.e., horizontal pairwise intra-speaker) match was used. For the Consonant lists, vertical pair match was provided with the pitch-loudness composite display. See BBN Report 2189 for fuller details on these algorithms, and the "Hints" booklet for examples drawn from this curriculum.

The first pass through the curriculum utilized the "A" versions of both the vowel and consonant lists. This was the only time they were used. The API software operated in a manner identical to that described in the above references.

TABLE 2

Fall and Spring semester training schedule of students through the word lists and levels of display disclosure. See text.

Word List	Teacher	Type of Disclosure: Full	Level 1	Level 2
INT 1	DNK	1	18,32	
"	DWD	2	19,33	39
INT 2	DNK	4	34	41
"	DWD	3,5		
VOW 1a	DNK	6		
	DWD	7		
VOW 2a	DNK	8		
	DWD	9		
VOW 3a	DNK	10		
	DWD	11		
VOW 1b	DNK		20	42
	DWD		21	
VOW 2b	DNK		22	38
	DWD		23	
VOW 3b	DNK		24	
	DWD		25	40
CON 1a	DNK	12		
	DWD	13		
CON 2a	DNK	14		
	DWD	15		
CON 3a	DNK	16		
	DWD	17		
CON 1b	DNK		26	
	DWD		27	37
CON 2b	DNK		28	36
	DWD		29	
CON 3b	DNK		30	
	DWD		31	35

Underlined session numbers were omitted from the Spring semester schedule, for a total of 37 training sessions, five less than the Fall semester's 42.

The second and third passes through the curriculum involved display modifications of a type not previously described. The principal feature of the API is its provision of immediate visual feedback derived and displayed in a manner that is relevant to the pronunciation parameter being trained. However, the very immediacy of that feedback may eventually become counterproductive if the student becomes overdependent on it for proper pronunciation. This is another way of stating the problem of generalization to normal, running speech. When outside the API situation, will students apply what they have learned to their everyday pronunciation? In an effort to facilitate this transfer, the immediacy of the visual feedback was reduced in two stages.

The second pass through the curriculum -- utilizing the "scrambled" versions of the vowel and consonant lists -- was done using "first-level delayed disclosure" of the display. Under the control of the switch register of the computer, the software operated normally save for one difference: when the student was using the STORE feature, the display of his analyzed speech did not appear, point-by-point, in real time as he spoke. It flashed on the screen immediately following the conclusion of his 2 1/2 second "time window." The teacher's display was disclosed in the normal fashion, as it was heard during both the STORE and REPLAY operations; similarly, the student's recording (placed on the tape loop in previous STOREs and seen at those times only in delayed mode) was disclosed gradually during the REPLAY process. The FREE mode, as always, provided instant analysis and display without the tape recording being made. The only delay, therefore, occurred when the student was speaking in STORE mode, and the delay always terminated 2 1/2 seconds from the start of that process. The student always had access to FREE mode when he wished immediate feedback.

The final pass -- the last eight sessions -- was done with "second-level delayed disclosure." Here again, immediate feedback was not provided in STORE mode; but the student's section of the display still remained blank at the conclusion of the entire STORE sequence. The student was forced to wait the additional 2 1/2 seconds during which the teacher's model utterance was heard and gradually disclosed; and even at the conclusion of this process, he still was not allowed to view the display produced by his speech. He had to actually make a response on the keyboard to indicate his readiness to view the output. Pressing the REPLAY button produced normal operation of that function: i.e., as the student section of the tape loop was heard, the display was disclosed. Pressing the MATCH button caused the display to appear and then move through the normal software manipulations. Pressing the FREE button produced normal operations, as before; and if the student entered FREE mode while silent, this also would cause the display of the previously STOREd utterance to appear, prior to being overwritten by speech later produced in FREE mode. Once the display of the previously-STOREd utterance had been enabled by any of the above means, the other functions also involving its display operated in their normal manner. Whenever the STORE mode was used again, its display was disabled until the appropriate extra response was made. In summary, then, the processes of delayed display served to lessen dependency on the visual feedback, and to allow the student to confirm his own ideas on the adequacy of his efforts to imitate the teacher, prior to seeing the API analysis. The "weaning" process was obviously incomplete, because immediate feedback was always available in free mode and because of the limitations of the curriculum material; but the students reported informally that they enjoyed the greater challenge of the delayed display.

Further reactions of some of the experimental students are contained in Appendix 5. These were solicited by the field engineer, towards the conclusion of their training.

2.5 Post- and Retention-testing

All students were retested with both the Lado test and the script-reading procedures. The post-test was administered to both groups within a 5-day period centering slightly later than the conclusion of the experimentals' training period. The retention-test was similarly tied to a point four weeks later, with a vacation intervening, and just prior to the final examinations in the Intensive English course.

One experimental student from the 12 in the Fall semester was forced to withdraw before the post-test; the rest completed all phases of training and testing. One control student in the fall semester withdrew from the Intensive English program before he could be post-tested; an additional two control students withdrew before they could be retention-tested. There were no dropouts in the Spring semester's students. In no case were the withdrawals related to student dissatisfaction with the API experiment.

Retention-testing was the final contact with the groups. The two types of information gathered from the students, over time, were then compared within and between students and treatment groups. It was naturally expected that individuals' performance would improve through time; what was to be tested was the hypothesis that the experimentals' scores would be increased for reasons traceable to their exposure to the API. The Lado test scores were directly analyzeable, and further discussion of those data will be deferred until Section 3 below.

The speech samples are far less tractable. In fact, what amounts to a new experiment must be done to extract the desired information from the data. As characterized in previous research of this type, the purpose of the experiment is to determine if the "signal" of training can be observed within the normal "noise" of the speech-production process. A judgment procedure is devised, and the speech behaviors produced by the groups through time are processed. The resultant numbers are analyzed with an eye to evaluating any pronunciation improvements peculiar to the experimental group. The most basic test that the data must pass is a subjective judgment process: one conducted using actual listeners and involving comparative (intra-speaker) judgments.

2.6 Preparation of Judgment Tapes

For each student, a composite tape was prepared containing his various attempts at certain test utterances across the three testing days. Each tape was organized as follows. There were three sections: intonation, vowels, and consonants. Within each section, utterances of a given type were randomized across the set of utterances and test days. Table 3 gives the utterances that were selected for subjective evaluation from the test-day script. The total number of utterances spliced into each judgment tape was 21 items times 3 testing days, or 63 per subject.

Each judgment tape began with the 18 utterances of the intonation set. The order of these utterances had been determined through the use of a deck of 18 cards, each calling for a specific utterance number and test day. The only constraint placed on the outcome of each shuffle of this deck was that a given utterance could not appear three times in succession. A different order was used for each subject. Five seconds' silence separated the

TABLE 3

Utterances Selected from Test Day Readings for Subjective Judgment

WORD GROUP 1: Intonation Phrases and Sentences

1. Is that a door?
2. What's that?
3. It's a map.
4. No, it's not a pen.
5. I saw Bill and Jane.
6. fruit basket

WORD GROUP 2: Vowel Contrasts

7.	deep	dip
8.	peep	pep
9.	luke	look
10.	cot	cut
11.	bake	back
12.	safe	surf
13.	loss	lice
14.	pot	pout
15.	seal	soil

WORD GROUP 3: Consonant Contrasts

16.	dare	tear
17.	gape	cape
18.	bat	vat
19.	shin	chin
20.	lacy	lazy
21.	do	threw

utterances within the group of 18. The final two sections of each judgment tape, 27 and 18 members respectively in length, utilized the same inter-item separation, and were delimited one from another by longer silent periods.

2.7 Accent-rating Procedures

Two other types of materials were produced for use in the accent-rating experiment. First, answer sheets for the judges were generated by a computer program which was fed both the actual utterances and the individual orders used in each judgment tape. The answer sheets, one set per judgment tape, were duplicated and used by each judge. Appendix 6 contains a complete set of three, covering one subject's tape. The purpose of the answer sheets was to inform the judges as to which utterance to expect next, to enable them to preset any internal criteria they might wish to muster. It also served to enable judges to respond in cases where gross pronunciation error might make utterance identification difficult or problematical. It contained no information about testing days on which the utterances were produced. Since the same recording conditions were used throughout, there were no differential cues of loudness or other artifacts to distinguish one set of utterance times from another.

Finally, sets of instructions were given to each judge, specifying the nature of the problem and the intervals of the scale they were to apply to the utterances. These instructions are included as Appendix 7 of this report. Each utterance was to be evaluated, in isolation, against a five-point scale of "apparent fluency," with higher scores being assigned to utterances as their fluency increased. An inspection of the actual instructions is the best way for the reader to familiarize himself with the

specifics of the scale that the judges were to use. Aside from the scale itself, judges were assured that it was only natural for adaptation effects to be noticeable, and that their confidence in their own ratings might increase through time.

The judgment process took a total of nine hours, distributed in three sessions on three successive days. Five adult BBN employees, native speakers of American English, served as judges. The 38 judgment tapes were played, in a random order over a loudspeaker at a comfortable listening level, in a relatively quiet conference room.

3. RESULTS

This section is divided into two parts, each dealing with one of the types of performance measurement employed to evaluate the effects of exposure to the API on accent-related variables.

3.1 Discrimination Test Data

The major question to be asked is this: is there a difference between the scores the students obtain as a result of their treatment? It is expected that students' scores will improve through time, due to the joint effects of familiarity with the test materials and course work. If, however, a reliable differential improvement in discrimination score could be ascribed to the experimental treatment, this would be a significant finding.

Table 4 summarizes the Lado Test scores for all students. The mean score for each testing day is shown below each column. Within each group, the ratios between these mean scores for the three pairwise comparisons between the test days are also shown.

A consideration of the distribution of data on this test, and of the documentation provided with the materials, makes it clear that scores are certainly not to be considered as arising from an underlying ratio scale. Due to ceiling effects, even some criteria for an interval scale are questionable here. It is therefore debatable whether a ratio of two test scores is meaningful in the manner implied by its computation. The answer is that of course the notion of "percentage improvement" should only be taken in a qualitative sense, as a rough yardstick indicating whether there is any point in delving further into the data with statistical tools requiring fewer assumptions.

TABLE 4
Complete Lado Test Data

	PRE	POST	RETN		PRE	POST	RETN
E1. Pomponio, I.	38	73	65	C1. Lopez, L.	34	44	56
E2. Gongora, A.	44	65	67	C2. Vegas, M.	45	54	50
E3. Mora, M.	37	58	66	C3. Luz, M.	32	54	60
E4. Peralta, V.	56	67	62	C4. Zuluaga, A.	51	62	66
E5. Bild, A.	58	66	72	C5. Szwarl, B.	59	78	74
E6. Cespedes, M.	58	75	70	C6. Vivas, S.	57	60	68
E7. Morelli, B.	69	76	79	C7. Franco, G.	54	54	57
E8. Demoya, T.	68	80	75	C8. Bermudez, I.	68	66	68
E9. Gomez, J.	58	68	71	C9. Orellana, R.	70	70	59
E10. Vegas, C.	71	81	77	C10. Priewer, C.	72	84	85
E11. Hernandez, C.	76	76	83	C11. Castellanos, M.	73	83	88
E12. Farberoff, J.	83	92	86	C12. Alalu, F.	80	83	91
E13. Compan, A.	52	65	68	C13. Kaplan, D.	56	79	84
E14. Acosta, M.	53	73	81	C14. Ruiz, C.	59	58	70
E15. Bermudez, I	68	68	62	C15. Gonzalez, E.	70	76	76
E16. Abumohor, A.	71	82	84	C16. Lechuga, C.	65	80	83
E17. Puga, M.	78	87	77	C17. Restrepo, M.	72	78	86
E18. Fraynd, P.	83	86	92	C18. Cocchis, A.	77	73	75
E19. Castellanos, M.	88	88	88	C19. Steinworth, M.	85	91	91
Means:	63.6	75.1	75.0		62.0	69.8	75.0
Ratios: POST/PRE	1.18				1.13		
RETN/PRE	1.18				1.18		
RETN/POST	1.00				1.05		

With these caveats in mind, the ratios between group means on the three test days can be considered. The major impression to be gained is of a lack of difference traceable to treatment. Both groups' test scores improved 18% on the average between pretest and retention test. This was accomplished by the experimental group in one step, from pretest to post-test; there was no further average improvement during the retention interval. The controls' improvement was more gradual: they improved 13% during the time the experimentals were being trained, and an additional 5% during the retention interval.

The equality between the groups in average improvement between pretest and retention test makes it fruitless to test for the existence of any significant difference. Even if one wished to make the attempt to transform the test scores to account for the nonlinearity presumably obtaining between score and "true discrimination ability," the fact that the students' pretest scores matched so closely ensures the failure of that proposed approach.

The only remaining comparison that is of primary interest is the difference of 5% between the two groups in the ratio of their average post-test score to their average pre-test score. To investigate this, a Mann-Whitney U test was performed on the difference scores of each subject (the signed difference between the post-test and the pre-test scores). There was no significant difference between the groups. Since there cannot be a significant group difference between the difference scores obtained from a retention minus pretest comparison, and further since transitivity is a necessity here, it follows that the observed 5% mean improvement between retention and post-testing for the control group is insignificantly different from the 0% change seen in the experimental group. Overall, then, service as an experimental subject produces no reliable improvement in discrimination ability over that which is produced by parent course work and/or familiarity with the measurement instrument.

3.2 Accent Ratings of Student Speech

As in the analysis of the previous data, the central question to be answered here is whether the experimental group's proficiency in the target language after training is significantly higher than that of the control group. The behaviors tapped here are on the production rather than the perceptual side of language skill, and the derivation of numerical data from this behavior is the more difficult because the measurement of pronunciation skill is a subtler task than is the measurement of discrimination skill. Certain simplifying assumptions were made in the course of the analysis of the judments produced through the methods outlined in sections 2.6 and 2.7 above. These assumptions were directed towards the extraction of overall changes in pronunciation skill while minimizing the variance introduced in the speech production and subjective judgment processes.

Each of the five judges produced a rating of the accent of the 21 different test utterances spoken at three points in time, by the 38 different subjects. The measurement of the changes in individual words or in individual subjects is not of central interest here; rather, the major variables to be addressed are whether the experimental students, as a group, improved significantly more than their control counterparts, for all the utterances tested. By extension, it is of interest to determine whether any differential improvement occurrs as a function of the word group involved. If this is observed, it will be relevant to the evaluation of the relative efficacy of the various displays and curricula used in attempting to improve certain aspects of accent.

A judgment is defined as the difference between the ratings assigned by one judge to one word spoken by one subject on two test days. Comparing, for example, the judgments given to a word spoken at pre- and post-testing times, a subject could receive a higher post rating, a lower post rating, or the same rating. If he received a higher post rating, the judgment is scored as an improvement in pronunciation of that word from the pretest to the post test, according to that judge. As the numbers of words, subjects, and judges increase, the proportions of occurrence of the three possible outcomes of such comparisons become amenable to statistical analysis. Two comparisons were made: pre vs post test and pre vs retention test. The results for the entire set of test words are presented first. Distinctions between performance on individual word groups will be dealt with next.

For the pre vs post test comparison, a large proportion of the judgments indicated a change in pronunciation ability. For the experimental group, 59 percent, and for the controls, 53 percent of all ratings given to post-test utterances were different than those given pre-test utterances of the same words. Within those changed ratings, 77 percent of the experimental group's went in the direction of improvement, while 62 percent of the controls' judgments were improved. The difference between these rates of improvement is significant (Chi-Square statistic $p<.001$, when computed on the 2 x 2 table including only changed judgments). The Chi-Square statistic computed on all pre-post judgments, when such judgments are dichotomized either as "improved" or as "no change or poorer," is also significantly different from chance. Table 5 gives more detail.

TABLE 5

Pre-Post Test Comparisons over All Words

	Experimentals		Controls	
	#	%	#	%
Total number of judgments indicating change	1174	59	1067	53
Total number of judgments indicating improvement	899	45	664	33
Total number of judgments indicating poorer pronunciation	275	14	403	20

X^2 including only judgments indicating change = 54.51
df = 1
p < .001

ϕ = .156

r = .243

X^2 including all judgments, dichotomized as either: (A) improved or (B) no change or poorer = 58.09
df = 1
p < .001

ϕ = .121

r = .189

Two additional statistics computed at the same time as the Chi-Square were the Phi Coefficient and the equivalent correlation coefficient, r. This latter figure gives a rough measure of the strength of the effect observed. The Phi coefficient is an underestimate of the actual correlation; Wert et. al. (1954) give tables and a rationale for the conversion of Phi values to correlation coefficients. The values of converted Phi coefficients range from 0 to 1, and can be interpreted in the same manner as standard r values. The significance levels of r values computed in this way must be evaluated in terms of the Chi-Square statistic, but the absolute level of the r statistic may be used as an estimate of the strength of the correlation between cells in the originating table. The equivalent correlation coefficients observed for the overall pre-post comparison were .24 for the changed judgments alone and .19 for all judgments.

The judgments made in the comparison of pretest vs retention-test utterances are similarly distributed. There was little change from the previous comparison in the distribution of the three types of judgment for the control group over all words, while the experimental group's responses contained more judgments that changed (63 percent here versus 59 percent for the pre-post comparison). Considering only judgments indicating change, 73 percent of the experimental group's changed judgments were in the direction of improvement, while 62 percent of the control group's pairs of changed ratings were judged as having improved in this pre-retention comparison. Two Chi-Square statistics computed similarly to those presented above indicated that the observed differences in these rates are significant ($p<.001$), though the equivalent r values are low -- .11 and .12 for the two castings of the contingency table. Table 6 gives the details.

TABLE 6

Pre-Retention Test Comparisons over All Words

	Experimentals		Controls	
	#	%	#	%
Total number of judgments indicating change	1255	63	1074	54
Total number of judgments indicating improvement	914	46	670	34
Total number of judgments indicating poorer pronunciation	341	17	404	20

X^2 including only judgments indicating change = 29.02
dF = 1
p < .001

ϕ = .112
r = .175

X^2 including all judgments, dichotomized as either:
(A) improved or (B) no change or poorer = 62.33
dF = 1
p < .001

ϕ = .125
r = .195

Before inspection of the differences between the two treatment groups on specific word groups, the numbers of judgments indicating "change" must be inspected in each of the three word groups. It has been shown above that the overall proportion of "change" judgments, for both the pre vs post-test and the pre vs retention-test comparisons, is lower for the control group. This difference in the relative proportion of "change" judgments holds uniformly throughout the three word groups, for both the pre-post and the pre-retention comparisons. Table 7 shows these data. By themselves, the differences in these proportions say nothing about the relative improvement of the two groups of subjects for these word groups, just as a comparison between the 59 percent and the 53 percent of "change" judgments for the overall pre-post comparison in Table 5 means little save in conjunction with the statistics involving the relative numbers of actual improvement judgments. The purpose of Table 6 is to assure the reader that, after this rough baseline of "change" responses is taken into account, meaningful comparisons may yet be made within these judgments, and that such comparisons may be made within the different word groupings. Given the parity between the results of the Chi-Square statistic when computed over the two dichotomizations presented in Tables 5 and 6, it is unnecessary to compute two such statistics below.

Tables 8 and 9 illustrate the differences between the two treatment groups, on each of the three types of curriculum material, for the pre-post and pre-retention comparisons. For the first word group, in which the judges rated the accentedness of phrases and sentences similar to the training materials in the intonation lists, the experimental subjects performed significantly ($p<.005$) better than the controls in both pre-post and pre-retention comparisons. Correlation coefficients were .17 and .27, respectively.

TABLE 7

Distribution of "Change" Judgments

Pre-Post Comparisons

	Word Group 1		Word Group 2		Word Group 3	
	# change judgments	% of all judgments	# change judgments	% of all judgments	# change judgments	% of all judgments
experimentals	366	64	505	59	303	53
controls	339	59	442	52	286	50
all subjects	705	62	947	55	589	52

Pre-Retention Comparisons

	Word Group 1		Word Group 2		Word Group 3	
	# change judgments	% of all judgments	# change judgments	% of all judgements	# change judgments	% of all judgments
experimentals	386	68	538	63	331	58
controls	338	59	467	55	269	47
all subjects	724	64	1005	59	600	53

TABLE 8

Pre- Post Test Comparisons by Word Group

Word Group 1: INTONATION, RHYTHM, + STRESS

	Experimentals	Controls
Total number of judgments indicating improvement	280	225
Total number of judgments indicating poorer pronunciation	86	114
	$X^2 = 8.89$ df = 1	$p < .005$
	$\phi = .112$	
	$r = .175$	

Word Group 2: VOWELS

	Experimentals	Controls
Total number of judgments indicating improvement	389	256
Total number of judgments indicating poorer pronunciation	116	186
	$X^2 = 39.63$ df = 1	$p < .001$
	$\phi = .204$	
	$r = .315$	

Word Group 3: CONSONANTS

	Experimentals	Controls
Total number of judgments indicating improvement	230	183
Total number of judgments indicating poorer pronunciation	73	103
	$X^2 = 9.98$ df = 1	$p < .005$
	$\phi = .130$	
	$r = .203$	

TABLE 9

Pre Retention Test Comparison by Word Group

Word Group 1: INTONATION, RHYTHM, + STRESS

	Experimentals	Controls
Total number of judgments indicating improvement	303	211
Total number of judgments indicating poorer pronunciation	83	127

X^2 = 22.6 df = 1 p < .001
ø = .177
r = .274

Word Group 2: VOWELS

	Experimentals	Controls
Total number of judgments indicating improvement	381	272
Total number of judgments indicating poorer pronunciation	157	195

X^2 = 17.37 df = 1 p < .001
ø = .131
r = .204

Word Group 3: CONSONANTS

	Experimentals	Controls
Total number of judgments indicating improvement	230	187
Total number of judgments indicating poorer pronunciation	101	82

X^2 = 0.0 df = 1
ø = 0
r = 0

The second word group tested the improvements in vowel pronunciation. Here, the largest individual word-group treatment effect was obtained in the pre-post comparison. The advantage of the experimental group was significant ($p<.001$) for both the pre-post and the pre-retention comparisons, but the strength of the relationship was larger ($r = .31$) for the pre-post than for the pre-retention comparison ($r = .20$).

The final word group, in which some consonant contrasts troublesome for Spanish-English speakers were tested, showed mixed results depending on the testing times. For the pre-post comparison, experimental subjects' utterances were significantly ($p<.005$) more likely to receive a judgment of improvement than the controls' utterances, and the equivalent r was .20; however, the advantage enjoyed by the experimentals disappears in the comparison of utterances made at pre versus retention-testing times, when both groups improved equally strongly.

4. DISCUSSION

In comparison with the results obtained in the smaller-scale experiment in the English-Mandarin Chinese language pair, the present data indicate greater success in altering the production of utterances calculated to emphasize the major components of the Spanish accents of the experimental students. However, this success seems of less consequence when considered in the light of the considerably greater exposure to the API system enjoyed by the experimental students in the present experiment. A further disappointment is contained in the discrimination data, which show no significant difference between the two treatment groups in the acquisition of English sound distinctions that are supposedly troublesome for Latin-American students.

The differences in performance observed between the three word groups, while suggestive, do not appear reliable. If any word group's performance is to be faulted, then it is the third, having to do with consonant contrasts. The curriculum here, and the displays produced, were among the most subtle and difficult to interpret; hence it is not surprising that the small differential improvement observed in the experimental group at the pre-post comparison should be lost by the end of the retention interval. The Vowel display, tested in word group 2, showed the strongest overall performance: the pre-post comparison's r value for the treatment effect reached its maximum value here for the entire experiment. This display was among the easiest to interpret, once the student picked up the correlation between tongue movement and display height. The intonation display, tested in word group 1, showed an intermediate level of efficacy. For this work, the pitch contours plotted on the API screen were immediately controllable by the students, and hence easily understood; but the length

of the training utterances and the need for consideration of a more complex pattern probably mitigated the strong treatment effect that might have been seen with shorter utterances. Of course, such are not germane to the present language pair, where the function of pitch contours, rhythm, and stress is at the suprasegmental level. While it is impossible to directly compare results across the two experiments conducted with the Mark II API, it was possible to demonstrate reliable differential training effects for pitch control at the segmental level for the English-Mandarin Chinese language pair.

At the same time that we state that the differential training effects were statistically reliable, for both the above experiments, we should also state that these effects are not considerable. The magnitude of the equivalent correlation coefficients does not indicate a very strong effect at all; indeed, one is led to the question of whether the reliability of the effect was demonstrated only because of the large numbers of words, subjects, and judges whose data were pooled to produce the ultimate 2 x 2 tables.

The narrowness of the evaluation procedures bears some additional discussion. The reader may have questioned an approach that limits the test day materials to written speech samples approximating the training curriculum, and that limits the attention of the accent-rating panel only to those aspects of the speech samples that relate to the display algorithms used by the experimental students. If the present evaluation procedures had shown unequivocal and strong advantages accruing to the experimental students, this approach would perhaps have been too constricted to

demonstrate its effects in everyday speech produced by the students. However, this design was planned with the realization that exactly such narrow training effects had the best chance of being produced, and also the best chance of being observed in a judgment of accent. It was felt that only if such narrow tests of the API's efficacy were passed did it become realistic to inspect more global aspects of speech behavior. It might have been more relevant to the evaluation of overall speech patterns if the students had produced speech extemporaneously on the test days, and if the accent-rating judges had been able to rate this speech in a more unstructured way. However, we judged that the variability and lack of focus of this approach would serve the ultimate purpose of the work less than the approach which was ultimately adopted. The orthographic confusions faced by the students in reading the test materials, the attentional demands made on the accent-rating panel, and the coarse grain of the numerical analysis procedure applied to the ratings were all prices that were paid in the hope that strong specific effects on separate parameters of accent could be produced and measured. The strength of the effects observed was such that one would not expect to find general improvements in English speech. Corroborating this impression are the data of the Lado discrimination test, which show no treatment effect. This test was chosen for the purpose of determining the overall efficacy of API exposure in improving English speech perception. Presumably, increased skill in the perceptive process would be reflected in speech production improvements as well. Both the treatment groups improved an equal amount throughout the experiment.

A consideration of the control treatment employed here also weakens our opinion of the strength of the observed differences between the two groups of students. API exposure was an addition to normal student work; the control group had had no supplementary pronunciation drill work that might have lessened its contrast with the experimental group, had that group's performance been greatly improved. The smallness of the observed difference between the treatments is further diminished by the lack of conservatism that circumstances forced in the specified control treatment.

The outcome of this program of research and development may be summarized in two statements:

(1) It is possible to demonstrate increased competence in the pronunciation of second languages in students who have used the API system.

(2) The observed changes do not appear to be cost-effective when compared to available alternatives.

The concept of improving the standard language laboratory paradigm through the addition of computer-implemented real-time analysis and visual display of speech parameters has thus been demonstrated in a limited way. The possible reasons for such limitations are several, and are interdependent.

First: the intermittent nature of the reinforcement given the students. The final experiment in this series afforded the largest opportunity for student exposure to the API system, and the results do not seem to be greatly different (in terms of significance) from those of the English-Chinese experiment. Perhaps not even 45 minutes per day of accent-reduction instruction

is able to affect speech behaviors so basically overlearned. An attempt was made to gradually reduce the immediacy of the visual feedback to the students, through the two-level disclosure scheme. However, this was an insufficient approximation to the ideal situation of a wearable pronunciation aid, whereby students could receive continuous articulatory feedback. The design and development of such a system is obviously many years in the future, though progress has been made on that front in the area of wearable speech-analyzing aids for hearing-impaired speakers.

Second: the limited nature of the training materials that were used. It would have been pedagogically preferable to have utilized a broader set of curriculum items, but the obvious limitations on system storage and display subtlety forced the materials into rather a narrow compass. This limitation was imposed as the consequence of a decision on design philosophy, discussed next.

Third: the decision to apply this analysis and display technology in an automated instructional environment, rather than as a system for the enhancement of the teacher-student interaction for pronunciation improvement. These two possible uses imply very divergent curricula and activities on the part of the student. Very early in this research, the present project's goal was determined to be the addition of visual feedback to the standard language laboratory situation. Students were to work alone, with prerecorded teacher materials. The absence of a live teacher (who presumably might be trained in the interpretation and manipulation of the displays) to work with the student necessitated the simplification of the display and curriculum to the point at which a relatively untrained (and certainly phonologically ignorant) student could extract useful and consistent information concerning the adequacy of his speech.

It has now been shown that it is infeasible to expose students to a limited curriculum focussing on one accent problem at a time and to expect this exposure to affect everyday speech behavior. The API is simply not interactive enough. Students would have greatly benefited from the additional attention of a tutor to keep tabs on pronunciation parameters not currently being displayed. If performance on these other aspects were to wander, the tutor could catch this early and redirect the student's attention to the neglected parameter(s).

During negotiations with several language-training institutions concerning the possibility of conducting the API field trials at those locations, it was often speculated that the system might be more efficiently applied in the remediation of accent problems where the student had both the need, the time, the motivation, and the perceptual ability required to make the effort worthwhile. For such an application, it is reasonable to employ a more personalized approach. The best indication of the potential success of the tutorial mode of system utilization is in the reactions that have been obtained in the related area of the instruction of hearing-impaired children with a computer system similar in several respects to the API (Nickerson and Stevens, 1973). It remains for the future to determine if similar success may be attained in the tutorial application of API techniques to accent reduction in second-language learning.

In this connection, the comments of the Director of the University of Miami Intensive English Program are relevant. These are included as Appendix 8 of the present report.

The fourth and final factor to be considered as a contributing element to the observed lack of substantial benefit of the API system in pronunciation improvement is the subtle nature of accent measurement itself. The present experiment utilized quite simplistic speech materials and judgment techniques. These were chosen precisely because of their simplicity and relevance to the accent parameters trained. The observed strength of the effects on accent in these test materials could only have been weakened if more global speech samples had been collected from the two groups of students. Another way of making the same point is to state that while exposure to a specific display may measureably affect the production of specific speech sounds when the experimental student has had his attention drawn to the need for accurate production of test materials, anecdotal evidence from observation of the normal English speech of the same students shows little effects of the training. It is apparent that if any method of pronunciation improvement -- computerized or not -- were to be truly effective in ameliorating a student's speech, then subtle measurement of that speech would be unnecessary; approaches like that are adopted, as it were, by acclamation. Certain teachers, perhaps blessed with apt students but certainly in possession themselves of pedagogical insight, can regularly achieve impressive results in pronunciation training. We have not yet proven that automated, computer-assisted speech instruction can be brought to a similar level.

INTENSIVE ENGLISH PROGRAM

P. O. BOX 8005
UNIVERSITY OF MIAMI
CORAL GABLES, FLORIDA 33124

• • •

PROGRAM DATES
October 2, 1972 – January 26, 1973
January 29, 1973 – May 18, 1973
May 21, 1973 – August 30, 1973

• • •

COURSE:
An intensive fifteen-week course especially designed to improve the ability to speak, write and understand English. Students may register **during the first three weeks of each course. Enrollment will normally be closed at the end of the third week.** Classes meet four hours a day, five days a week for a period of fifteen weeks. For administrative reasons it is impossible to establish class schedules prior to the beginning of the course. Applicants must be at least seventeen years of age. Upon arrival, each student is tested, and on the basis of his test score, assigned to the Elementary, Intermediate or Advanced level of the Program.

ELEMENTARY: (Fifteen weeks) For the beginning student or the student whose knowledge of English is very limited; one hour of reading, one of grammar, one of conversation and an hour of laboratory instruction and practice in pronunciation and the differentiation of sounds in English.

INTERMEDIATE: (Fifteen weeks) Basic grammatical constructions, more intensive reading, sentence and paragraph construction, idioms of conversation and oral-aural laboratory training. Open to students who have completed Elementary or those with Intermediate Intensive English Program Placement Test results.

ADVANCED: (Fifteen weeks) Fine points of grammar, selected readings in prose and poetry, composition, discussion and study of idioms. A fifth hour of laboratory practice in pronunciation for five weeks, followed by lectures on United States history, geography and government. Open to students who have completed Intermediate or those with Advanced Intensive English Program Placement Test results. Four elective credits are awarded for each level of the program satisfactorily completed, to those students who remain at the University of Miami and enroll as regular degree-seeking students. The University accepts satisfactory completion of the Advanced level of the Intensive English Program as proof of English proficiency. However, admission to the Intensive English Program in no way guarantees acceptance to the University as a degree-seeking student.

Those wishing to remain as regular students after they finish their English study must make formal application and submit credentials as required by the Office of Admissions.

Counselors are available at all times to assist students with problems both academic and personal. An extra hour of classroom instruction may be required of students who have difficulty writing the Roman alphabet or who have special pronunciation problems. Only those students who attend the last twelve weeks, complete all course requirements, and pass final examinations are awarded *Certificates.*

ACTIVITIES:
Intensive English students are included in University-wide activities. In addition, special events are scheduled and invitations to visit North American homes are extended to better acquaint the students with the United States and its citizens.

UNIVERSITY:
The University of Miami has a large modern campus located a short distance from downtown Coral Gables. It boasts many new buildings including a Student Union with a cafeteria, coffee shop, ping-pong, billiards and bowling rooms and an Olympic-size swimming pool. There are also many attractive classroom buildings and a completely air-conditioned library.

COMMUNITY:
The University of Miami is located in suburban Coral Gables, a residential community adjacent to Miami in the heart of the Gold Coast of South Florida, just 15-20 minutes from the Miami International Airport. There is bus service between Coral Gables, Miami and Miami Beach. During vacation periods, many students plan sightseeing trips to other North American cities.

CLIMATE:
A subtropical climate prevails, and average temperatures range from 60°F. to 80°F. during most of the year, requiring only light summer-weight clothing, sweaters and raincoats.

FOR YOUR INFORMATION

1. The Intensive English Program is often referred to as IEP.

2. ABSENCES

   a. All students must attend all scheduled classes. Each student is allowed five (5) unexcused absences to take care of necessary business such as going to the bank, to the Immigration Office or to the airport.

   b. Excuses are given by the Intensive English Office only in cases of illness. A student who has been ill must report to the Office the day he or she returns to class and have the absence excused. (Report to Mrs. Brodigan.)

   c. Students having more than five unexcused absences will not be eligible for a Certificate. Students having excessive absences will be required to withdraw from the IEP course and will be out of status with the Office of Immigration and Naturalization. Students who are withdrawn may not live in University housing.

   d. Teachers do not excuse absences. However, if a student knows he will be absent, he should so inform his teacher. It is a student's responsibility to make up work he has missed because of his absence. Make-up tests must be arranged by the student with the teacher.

3. GRADES

   a. Interim grades are given at the end of five weeks of classes and at the end of ten weeks of classes. Final grades are given at the end of the semester. Grades are based on written work, oral tests, homework, and class participation. The following grading system is used:

      A - excellent
      B - above average
      C - average, or satisfactory
      D - below average
      E - unsatisfactory

   b. To pass the course and be eligible for a Certificate, a student must have a grade average of "C" and no grade of "E".

4. CERTIFICATES

   a. A Certificate is granted at the end of the semester to those students who meet all of the following four requirements:

      1) Take the final examination
      2) Pass the course with a "C" average or better
      3) Have satisfactory attendance
      4) Have paid all University fees in full

   b. Students who do not qualify for a Certificate receive a letter which explains the reason they failed to quality.

   c. Four hours of elective credits are awarded for each level of the program satisfactorily completed to those students who enroll as regular degree-seeking students at the University of Miami.

INTENSIVE ENGLISH PROGRAM
CE250 SPRING 1974

SCHEDULE

Mon. Jan. 7	9:00 A.M. Registration - Bldg. 48
Tues. Jan. 8	Classes begin
Wed. Jan. 9	Reception for students - Student Union S226, 7:30 - 9:30 P.M.
Fri. Feb. 8	End of first five weeks
Tues. Feb. 12	Interim grades due from teachers - 10:00 A.M.
Fri. Mar. 8	TOEFL
Sat. Mar. 9 to Sun. Mar. 24	Spring Recess
Fri. Mar. 29	End of 2nd five weeks
Tues. Apr. 2	Interim grades due from teachers - 10:00 A.M.
Mon., Tues., Wed., Apr. 29, 30 & May 1	Final Examinations
Thurs., May 2	Regular classes Final grades due from teachers - 9:00 A.M.
Fri., May 3	Graduation - Student Union S226 1:00 P.M. - 3:00 P.M.
Mon. May 6, 1974	Registration for Summer Semester - 9:00 A.M.

HOLIDAYS

IN THE UNITED STATES, SUNDAY IS THE ONLY HOLIDAY RECOGNIZED BY COMMON LAW. THERE ARE NO NATIONAL HOLIDAYS. EACH STATE HAS THE AUTHORITY TO DESIGNATE THE HOLIDAYS IT WILL OBSERVE. THE PRESIDENT ISSUES A PROCLAMATION SETTING ASIDE A SPECIAL HOLIDAY, BUT THAT PROCLAMATION IS MANDATORY ONLY FOR FEDERAL EMPLOYEES AND DISTRICT OF COLUMBIA. THE GOVERNOR OF EACH STATE ISSUES A SIMILAR PROCLAMATION FOR HIS STATE, USUALLY THE SAME AS THAT OF THE PRESIDENT. BUT NOT NECESSARILY.

CALENDAR

*indicates the days government offices, banks, schools, and most stores and businesses are closed.

*January 1 - New Year's Day

February 11 - Abraham Lincoln's Birthday

February 14 - Valentine's Day - cards (called Valentines), candy and flowers are given as a sign of love or friendship.

February 22 - George Washington's Birthday - George Washington was the first President of the United States.

April 1 - April Fool's Day - a day when children play tricks on each other or try to fool people, e.g., putting salt in the sugar bowl.

May - - 2nd Sunday in May - Mother's Day

*May - last Monday in May - Memorial Day - a day to pay respect to the dead.

June - 3rd Sunday in May - Father's Day

*July 4 - Independence Day

*September - 1st Monday - Labor Day

October - 2nd Monday in October - Columbus Day

October - 4th Monday in October - Veteran's Day - a day to honor all people who have served in the Armed Forces - banks are usually closed - some states designate Nov. 11.

October 31 - Halowe'en - originated from a time when people belived in witches, ghosts and evil spirits and thought that they could be scared away if people wore masks and costumes - children dressed in costumes and go from house to house with large sacks to carry the fruit and candy people give them - it is a time for pumpkins, apple cider, skeletons, and black cats.

*November - 4th Thursday in November - Thanksgiving Day - a day to give thanks to God for all the blessings of the past year - it is a family day, celebrated with big dinners and joyous reunions.

*December 25 - Christmas Day - a day celebrated in all Christian communities as the birth of Jesus Christ - it is a day to exchange gifts, common customs include sending greeting cards to friends; the Christmas tree, a gaily decorated evergreen tree; Santa Claus, a genial, jolly gentleman whose wife and elves spend the year making toys and then load them into Santa's sleigh and hitch eight reindeer to it so Santa can fly around the world on Christmas Eve delivering gifts; and singing traditional songs.

There are other religious celebrations such as Yom Kippur and Hannukah (Jewish) and Easter (Christian).

APPENDIX 2

Script Read by Experimental and Control Students on Testing Days.

Report No. 2841

50 Moulton Street
Cambridge, Mass. 02138
Telephone (617) 491-1850

Bolt Beranek and Newman Inc.

AUTOMATED PRONUNCIATION INSTRUCTOR SYSTEM EVALUATION PROJECT

BOLT BERANEK AND NEWMAN INC.

UNIVERSITY OF MIAMI, CORAL GABLES

INTENSIVE ENGLISH PROGRAM

SCRIPT

for

STUDENT VOICE SAMPLES

* * * * * * * * * * * * *

STUDENT: (Please read the following sentences and fill in your name and the date.)

"MY NAME IS _____."

"TODAY'S DATE IS _____."

(Now turn the page and continue reading as well as you can.)

1. What's this?

2. It's a window.

3. Is this a chair?

4. No, it's a desk.

5. Is this a pen?

6. No, it's a pencil.

1. Is that a door?
   10 2 20 15

2. No, it's a picture.
   11 10 20 10 3

3. What's that?
   19 2

4. It's a map.
   10 20 2

5. No, it's not a pen.
   11 10 12 20 6

6. Who am I?
   14 20 9

7. That's a chair.
   2 20 4

8. This is November.
   10 10 11 5 3

9. I saw Bill and Jane.
   9 15 10 20 1

10. Fruit basket
    14 2 10

1. Is he at ease?
   10 5 2 5

2. Tell them a funny tale.
   6 6 20 19 10 1

3. The fool is full of dinner.
   20 14 10 15 20 10 8

4. Did you cut your coat?
   10 14 19 15 11

5. The cat is on the cot.
   20 2 10 13 20 12

6. He is safe in the surf.
   5 10 1 10 20 8

7. Put that seed to one side.
   15 2 5 20 19 9

8. That lout got his loot.
   2 17 12 10 14

9. Boil it in a bowl.
   16 10 10 20 11

10. Did he dare to tear it?
    10 5 4 14 4 10

11. Pat him with the bat.
    2 10 10 20 2

12. He got the cot.
    5 12 20 12

13. The bat is in the vat.
    20 2 10 10 20 2

14. The mayor met the major.
    20 1 3 6 20 1 8

15. She broke her chin and her shin.
    5 11 8 10 2 8 10

16. Mr. Lacy is lazy.
    1-10 10 1-10

17. What did you do when he threw it?
    19 10 18 14 6 5 14 10

18. These letters are "d's"
    5 6 8 3 5

	1. BEET, BIT, BEET, BIT	$\bar{E}$ - I ii - ɪ 5 - 10
$\bar{E}$ - $\bar{A}$ ii - ei 5 - 1	2. SEE, SAY, SEE, SAY	
	3. DEEP, DIP, DEEP, DIP	$\bar{E}$ - I ii - ɪ 5 - 10
$\bar{A}$ - I ei - I 1 - 10	4. HATE, HIT, HATE, HIT	
	5. HATE, HEAD, HATE, HEAD	$\bar{A}$ - E ei - ɛ 1 - 6
$\bar{E}$ - E ii - ɛ 5 - 6	6. PEEP, PEP, PEEP, PEP	
	7. LJKE, LOOK, LJKE, LOOK	$\bar{O}\bar{O}$ - 00 uu - ʊ 14 - 15
$\bar{O}$ - 00 ou - ʊ 11 - 15	8. GOAD, GOOD, GOAD, GOOD	
	9. COAT, CUT, COAT, CUT	$\bar{O}$ - U ou - ʌ 11 - 19
O - U ɑ - ʌ 12 - 19	10. COT, CUT, COT, CUT	
	11. POT, PAT, POT, PAT	O - A ɑ - ae 12 - 2
$\bar{A}$ - A ei - ae 1 - 2	12. BAKE, BACK, BAKE, BACK	
	13. BONE, BURN, BONE, BURN	$\bar{O}$ - ($\widetilde{ER}$ - $\widetilde{UR}$) ou - ɚ 11 - 8
$\bar{A}$ - ($\widetilde{ER}$ - $\widetilde{UR}$) ei - ɚ 1 - 8	14. SAFE, SURF, SAFE, SURF	
	15. HEAT, HATE, HEAT, HATE	$\bar{E}$ - $\bar{A}$ ii - ei 5 - 1
$\bar{O}\bar{O}$ - $\bar{O}$ uu - ou 14 - 11	16. SUE, SO, SUE, SO	
	17. NIECE, NICE, NIECE, NICE	$\bar{E}$ - $\bar{I}$ ii - aɪ 5 - 9

	18. LOSS, LICE, LOSS, LICE		0 - $\bar{I}$ ɑ - aɨ 12 - 9
$\bar{OO}$ - OU uu - au 14 - 17	19. LOOT, LOUT, LOOT, LOUT		
	20. POT, POUT, POT, POUT		0 - OU ɑ - au 12 - 17
$\bar{O}$ - OI ou - ØI 11 - 16	21. LOAN, LOIN, LOAN, LOIN		
	22. SEAL, SOIL, SEAL, SOIL		$\bar{E}$ - OI ii - ØI 5 - 16
D - T	23. DARE, TEAR, DARE, TEAR		
	24. BAT, PAT, BAT, PAT	B - P	
G - K	25. GAPE, CAPE, GAPE, CAPE		
	26. BAT, VAT, BAT, VAT	B - V	
J - Y	27. MAJOR, MAYOR, MAJOR, MAYOR		
	28. SHIN, CHIN, SHIN, CHIN	SH - CH	
S - Z	29. LACY, LAZY, LACY, LAZY		
	30. DO, THREW, DO, THREW	D - TH (voiceless)	
D - TH (voiced)	31. DIE, THY, DIE, THY		

APPENDIX 3

Curriculum Listings

LIST CODE INT #1 TITLE: Intonation contours: Phrases and sentences

DISCRIMINATION #	UTTERANCE	FUNCTION DISPLAYED
	1. What's this?	PITCH or PITCH-LOUDNESS
	2. It's nine o'clock.	
	3. She's a nurse.	
	4. That's a book.	
	5. What is your name?	
	6. Today is Monday.	
	7. airplane	
	8. How are you?	
	9. Good morning.	
	10. Where are you going?	
	11. This is a classroom.	
	12. Breakfast is ready.	
	13. I'm a student.	
	14. railroad station	
	15. telephone booth	
	16. traffic sign	
	17. That's a new pencil.	
	18. I see four airplanes.	
	19. I bought a suit.	
	20. I adore ice cream.	
	21. Do you have a book?	
	22. May I help you?	
	23. Is he sleeping?	
	24. Is this an apple?	

LIST CODE INT #2

TITLE: Intonation contours: Sentences
(From Wright-McGillivray Units 2, 3)

DISCRIMINATION #	UTTERANCE	FUNCTION DISPLAYED
1.	Is this a classroom?	PITCH or PITCH-LOUDNESS
2.	Yes, It's a classroom.	
3.	Is this a floor?	
4.	Yes, it's a floor.	
5.	Is this a ruler?	
6.	Yes, it's a ruler.	
7.	Is that a light?	
8.	Yes, that's a light.	
9.	No, it's not a door.	
10.	It's a window.	
11.	No, it's not a map.	
12.	It's a picture.	
13.	What are these?	
14.	They're keys.	
15.	What are those?	
16.	They're words.	
17.	They're sentences.	
18.	Are those desks?	
19.	No, they're not desks.	
20.	Is there a map here?	
21.	Where is it?	
22.	It's on the desk.	
23.	Is it here or there?	
24.	Is it blue or green?	

LIST CODE VOW #1a TITLE: Vowels 1: Tense/lax, straight version

DISCRIMINATION	#	UTTERANCE			FUNCTION DISPLAYED
/i/ - /I/	1.	bead	bid	$\bar{E}$ - I ii - ɪ	EI
	2.	beet	bit	5 - 10	
	3.	seek	sick		
/e/ - /I/	4.	bade	bid	$\bar{A}$ - I ei - I	EI
	5.	bait	bit	1 - 10	
	6.	cake	kick		
/e/ - /ɛ/	7.	bade	bed	$\bar{A}$ - E ei - ɛ	EI
	8.	bait	bet	1 - 6	
	9.	fade	fed		
/i/ - /ɛ/	10.	bead	bed	$\bar{E}$ - E ii - ɛ	HL
	11.	beet	bet	5 - 6	
	12.	feed	fed		
/u/ - /U/	13.	suit	soot	$\overline{OO}$ - OO uu - ʊ	OO
	14.	cooed	could	14 - 15	
	15.	wooed	wood		
/o/ - /U/	16.	coke	cook	$\bar{O}$ - OO ou - ʊ	FB
	17.	code	could	11 - 15	
	18.	boat	book		
/o/ - /ʌ/	19.	boat	but	$\bar{O}$ - U ou - ʌ	HL
	20.	bode	bud	11 - 19	
	21.	soap	sup		
/a/ - /ʌ/	22.	hot	hut	O - U ɑ - ʌ	AA
	23.	sod	sud	12 - 19	
	24.	dog	dug		

LIST CODE VOW #1b

TITLE: Vowels 1: Tense/lax, scrambled version

DISCRIMINATION	#	UTTERANCE		FUNCTION DISPLAYED
/i/ - /I/	1.	bead	bid	EI
	2.	bit	beet	
	3.	seek	sick	
/e/ - /I/	4.	bid	bade	EI
	5.	bait	bit	
	6.	kick	cake	
/e/ - /ɛ/	7.	bade	bed	EI
	8.	bet	bait	
	9.	fade	fed	
/i/ - /ɛ/	10.	bed	bead	HL
	11.	beet	bet	
	12.	fed	feed	
/u/ - /ʊ/	13.	suit	soot	OO
	14.	could	cooed	
	15.	wooed	wood	
/o/ - /ʊ/	16.	cook	coke	FB
	17.	code	could	
	18.	book	boat	
/o/ - /ʌ/	19.	boat	but	HL
	20.	bud	bode	
	21.	soap	sup	
/a/ - /ʌ/	22.	hut	hot	AA
	23.	sod	sud	
	24.	dug	dog	

LIST CODE VOW #2a TITLE: Vowels 2: Low/mid, tense/lax II straight version

DISCRIMINATION	#	UTTERANCE		FUNCTION DISPLAYED
/a/ - /æ/	1. sop	sap	O - A æ - ae	AA
	2. hot	hat	12 - 2	
	3. sod	sad		
/e/ - /æ/	4. fade	fad	Ā - A ei - ae	HL
	5. base	bass	1 - 2	
	6. bait	bat		
/o/ - /ɚ/	7. so	sir	Ō - (ẽR - ũR) ou - ɚr	FB
	8. code	curd	11 - 8	
	9. boat	bert		
/e/ - /ɚ/	10. hate	hurt	Ā - (ẽR - ũR) ei - ɚr	ER
	11. bait	bert	1 - 8	
	12. bade	bird		
/i/ - /I/	13. bead	bid	Ē - I ii - ɪ	EI
	14. beet	bit	5 - 10	
	15. seek	sick		
/e/ - /ɛ/	16. bade	bed	Ā - E ei - ɛ	EI
	17. bait	bet	1 - 6	
	18. fade	fed		
/u/ - /Ü/	19. suit	soot	ŌŌ - OO uu - ʊ	OO
	20. cooed	could	14 - 15	
	21. wooed	wood		
/o/ - /ʌ/	22. boat	but	Ō - U ou - ʌ	HL
	23. bode	bud	11 - 19	
	24. soap	sup		

LIST CODE VOW #2b TITLE: Vowels 2: Low/mid, tense/lax II
scrambled version

DISCRIMINATION	#	UTTERANCE		FUNCTION DISPLAYED	
/a/ - /æ/	1.	sop	sap	O - A æ - ae	AA
	2.	hat	hot	12 - 2	
	3.	sod	sad		
/e/ - /æ/	4.	fad	fade	Ā - A ei - ae	HL
	5.	base	bass	1 - 2	
	6.	bat	bait		
/o/ - /ɚ/	7.	so	sir	Ō - (ẼR - ŨR) ou - ɚ	FB
	8.	curd	code	11 - 8	
	9.	boat	bert		
/e/ - /ɚ/	10.	hurt	hate	Ā - (ẼR - ŨR) ei - ɚ	ER
	11.	bait	bert	1 - 8	
	12.	bird	bade		
/i/ - /I/	13.	bid	bead	Ē - I ii - ɪ	EI
	14.	beet	bit	5 - 10	
	15.	sick	seek		
/e/ - /ɛ/	16.	bed	bade	Ā - E ei - ɛ	EI
	17.	bait	bet	1 - 6	
	18.	fed	fade		
/u/ - /ʊ/	19.	soot	suit	ŌŌ - OO uu - ʊ	OO
	20.	cooed	could	14 - 15	
	21.	wood	wooed		
/o/ - /ʌ/	22.	but	boat	Ō - Ū ou - ʌ	HL
	23.	bode	bud	11 - 19	
	24.	sup	soap		

LIST CODE VOW #3a TITLE: Vowels 3: Diphthongs/diphthongized straight version

DISCRIMINATION	#	UTTERANCE		FUNCTION DISPLAYED
/i/ - /e/	1. beat	bait	Ē - Ā ii - ei 5 - 1	HL
	2. feed	fade		
	3. see	say		
/u/ - /o/	4. boot	boat	ŌŌ - Ō uu - ou 14 - 11	HL
	5. shoe	show		
	6. soup	soap		
/i/ - /ai/	7. heed	hide	Ē - Ī ii - aɨ 5 - 9	HL
	8. seat	sight		
	9. heat	height		
/a/ -/ai/	10. cot	kite	O - Ī ɑ - aɨ 12 - 9	HL
	11. sod	side		
	12. hot	height		
/u/ - /au/	13. who	how	ŌŌ - OU uu - au 14 - 17	HL
	14. shoot	shout		
	15. boot	bout		
/a/ - /au/	16. cod	cowed	O - OU ɑ - au 12 - 17	HL
	17. shot	shout		
	18. dot	doubt		
/o/ - /ɔi/	19. joe	joy	Ō - OI ou - ØI 11 - 16	FB
	20. so	soy		
	21. pose	poise		
/i/ - /ɔi/	22. bees	boys	Ē - OI ii - ØI 5 - 16	FB
	23. bead	boyd		
	24. see	soy		

LIST CODE VOW #3b TITLE: Vowels 3: Diphthongs/diphthongized scrambled version

DISCRIMINATION	#	UTTERANCE		FUNCTION DISPLAYED	
/i/ - /e/	1.	beat	bait	Ē - Ā ii - ei 5 - 1	
	2.	fade	feed		
	3.	see	say		
/u/ - /o/	4.	boat	boot	ŌŌ - Ō uu - ou 14 - 11	
	5.	shoe	show		
	6.	soap	soup		
/i/ - /ai/	7.	heed	hide	Ē - Ī ii - aɨ 5 - 9	
	8.	sight	seat		
	9.	heat	height		
/a/ - /ai/	10.	kite	cot	O - Ī ɑ - aɨ 12 - 9	
	11.	sod	side		
	12.	height	hot		
/u/ - /au/	13.	who	how	ŌŌ - OU uu - au 14 - 17	
	14.	shout	shoot		
	15.	boot	bout		
/a/ - /au/	16.	cowed	cod	O - OU ɑ - au 12 - 17	
	17.	shot	shout		
	18.	doubt	dot		
/o/ - /oi/	19.	joe	joy	Ō - OI ou - ØI 11 - 16	
	20.	soy	so		
	21.	pose	poise		
/i/ - /ɨi/	22.	boys	bees	Ē - OI ii - ØI 5 - 16	
	23.	bead	boyd		
	24.	soy	see		

LIST CODE CON #1a TITLE: Consonants: Aspirate/unaspirate initials straight version

DISCRIMINATION	#	UTTERANCE		FUNCTION DISPLAYED
/d/ - /t/	1.	deem	team	PITCH-LOUDNESS
	2.	dot	tot	
	3.	do	to	
	4.	dan	tan	
	5.	dame	tame	
	6.	doe	toe	
	7.	dip	tip	
	8.	dime	time	
/b/ - /p/	9.	beak	peak	
	10.	bin	pin	
	11.	bale	pail	
	12.	bet	pet	
	13.	bad	pad	
	14.	bought	pot	
	15.	bun	pun	
	16.	buy	pie	
/g/ - /k/	17.	gill	kill	
	18.	goat	coat	
	19.	gay	kay	
	20.	good	could	
	21.	gut	cut	
	22.	gab	cab	
	23.	gate	kate	
	24.	got	caught	

LIST CODE CON #1b TITLE: Consonants: Aspirate/unaspirate initials scrambled version

DISCRIMINATION	#	UTTERANCE		FUNCTION DISPLAYED
/d/ - /t/	1.	deem	team	PITCH-LOUDNESS
	2.	tot	dot	
	3.	do	to	
	4.	tan	dan	
	5.	dame	tame	
	6.	toe	doe	
	7.	dip	tip	
	8.	time	dime	
/b/ - /p/	9.	beak	peak	
	10.	pin	bin	
	11.	bale	pail	
	12.	pet	bet	
	13.	bad	pad	
	14.	pot	bought	
	15.	bun	pun	
	16.	pie	buy	
/g/ - /k/	17.	gill	kill	
	18.	coat	goat	
	19.	gay	kay	
	20.	could	good	
	21.	gut	cut	
	22.	cab	gab	
	23.	gate	kate	
	24.	caught	got	

LIST CODE CON #2a TITLE: Consonants 2: straight version

DISCRIMINATION	#	UTTERANCE		FUNCTION DISPLAYED
/b/ - /v/	1.	boat	vote	PITCH-LOUDNESS
	2.	buoy	very	
	3.	cupboard	covered	
	4.	rebel	revel	
	5.	harbored	harvard	
	6.	robe	rove	
	7.	curb	curve	
/j/ - /y/	8.	jeer	year	
	9.	jail	yale	
	10.	joke	yolk	
	11.	juice	use	
	12.	jet	yet	
	13.	jack	yak	
	14.	jello	yellow	
	15.	jewel	you'll	
	16.	jell	yell	
/š/ - /č/	17.	dish	ditch	SH - CH
	18.	crush	crutch	
	19.	wash	watch	
	20.	cash	catch	
	21.	wish	witch	
	22.	she's	cheese	
	23.	sheep	cheap	
	24.	shoes	choose	

LIST CODE CON #2b TITLE: Consonants 2: scrambled version

DISCRIMINATION	#	UTTERANCE		FUNCTION DISPLAYED
/b/ - /v/	1.	boat	vote	PITCH-LOUDNESS
	2.	very	bury	
	3.	cupboard	covered	
	4.	revel	rebel	
	5.	harbored	harvard	
	6.	rove	robe	
	7.	curb	curve	
/j/ - /y/	8.	year	jeer	
	9.	jail	yale	
	10.	yolk	joke	
	11.	juice	use	
	12.	yet	jet	
	13.	jack	yak	
	14.	yellow	jello	
	15.	jewel	you'll	
	16.	yell	jell	
/š/ - /č/	17.	dish	ditch	SH - CH
	18.	crutch	crush	
	19.	wash	watch	
	20.	catch	cash	
	21.	wish	witch	
	22.	cheese	she's	
	23.	sheep	cheap	
	24.	choose	shoes	

LIST CODE CON #3a TITLE: Consonants 3: straight version

DISCRIMINATION	#	UTTERANCE		FUNCTION DISPLAYED
/s/ - /z/	1.	sue	zoo	PITCH-LOUDNESS
	2.	sink	zinc	
	3.	race	raise	
	4.	ice	eyes	
	5.	bus	buzz	
	6.	niece	knees	
	7.	loose	lose	
	8.	advice	advise	
/d/ - /θ/	9.	dinner	thinner	D - TH (voiceless)
	10.	drill	thrill	
	11.	drew	threw	
	12.	dirty	thirty	
	13.	pad	path	
	14.	mad	math	
	15.	claude	cloth	
	16.	bread	breath	
/d/ - /ð/	17.	day	they	D - TH (voiced)
	18.	doze	those	
	19.	dough	though	
	20.	dare	their	
	21.	dave	they've	
	22.	ladder	lather	
	23.	fodder	father	
	24.	breed	breathe	

LIST CODE CON #3b TITLE: Consonants 3: scrambled version

DISCRIMINATION	#	UTTERANCE		FUNCTION DISPLAYED
/s/ - /z/	1.	sue	zoo	PITCH-LOUDNESS
	2.	zinc	sink	
	3.	race	raise	
	4.	eyes	ice	
	5.	bus	buzz	
	6.	knees	neice	
	7.	loose	lose	
	8.	advise	advice	
/d/ - /θ/	9.	dinner	thinner	D - TH (voiceless)
	10.	thrill	drill	
	11.	drew	threw	
	12.	thirty	dirty	
	13.	pad	path	
	14.	math	mad	
	15.	claude	cloth	
	16.	breath	bread	
/d/ - /ð/	17.	day	they	D - TH (voiced)
	18.	those	doze	
	19.	dough	though	
	20.	their	dare	
	21.	dave	they've	
	22.	lather	ladder	
	23.	fodder	father	
	24.	breathe	breed	

APPENDIX 4

Instructions for Experimental Students

Report No. 2841

50 Moulton Street
Cambridge, Mass. 02138
Telephone (617) 491-1850

Bolt Beranek and Newman Inc.

AUTOMATED PRONUNCIATION INSTRUCTOR SYSTEM EVALUATION PROJECT

BOLT BERANEK AND NEWMAN INC.

UNIVERSITY OF MIAMI, CORAL GABLES

INTENSIVE ENGLISH PROGRAM

HINTS FOR THE STUDENT

The purpose of this handout is to help you to understand what it is the computer is displaying to you and how it relates to your speech. Each of the various groups of words that you will be working with was selected to help you with a particular problem of speech that you might encounter in learning English. Similarly, the computer will produce a visual display of something that is relevant to that speech problem for you to use your eyes on, in addition to your ears as you normally do in listening to your own and the teacher's speech. What do we mean by "relevant"? We mean that the picture on the screen is responding to something in your speech that needs your full attention, while other aspects of your speech may be irrelevant. Consequently, those other aspects are not allowed to clutter up the picture that you see. We have tried to simplify the pictures so that they may be easily understood.

All of the pictures shown on the screen are visual descriptions of the time course of some aspect(s) of both your own and the teacher's (pre-recorded) speech. The bottom section of the screen reads from left to right as time proceeds, and shows something about your speech as time moves from the beginning to the end of your utterance. The same thing holds for the teacher model displayed in the upper part of the picture. Your task is to produce an utterance that both sounds as close to the teacher as possible, and also produces a visual display that looks as similar as possible to the teacher's visual display. Since your native language is not English, you may have trouble in hearing differences between your speech and the teacher's speech. At this point, we do not need to go into the reasons for that possible inability on your part. Let it only be said that an utterance of yours that you feel is very similar to the teacher's may not be judged so

by a native speaker of English. It takes a special ability, which some of you may possess, called "an ear for languages", to be able to modify your pronunciation in such a way that it is more similar to the teacher's, although it may then sound unnatural to you. We hope that the visual display provided by this system will help you in improving the naturalness of your English speech, and will allow even students without that "ear for languages" to reduce their accent in English.

Using the system for this purpose obviously involves understanding the visual display. To concentrate only on the tape recorded sound comparisons that the REPLAY button allows is to avoid most of the benefits that the system can provide. Intelligent use of the visual display requires some basic understanding of the nature of the display, and some specific knowledge about what aspect of speech is being shown for each of the word lists. Furthermore, some experience or knowledge is needed to be able to discriminate the important from the irrelevant aspects of the pictures that are shown. The main purpose of this document is to give you some ideas and pointers that you may use in trying to understand when the pictures indicate that your performance is acceptable, certain types of common errors to look for, the ways to correct them, and parts of the display that are not reliable or irrelevant to a good accent for that particular aspect of speech. Please note that the system will never explicitly tell you, via sight or sound, that your pronunciation is correct or incorrect. It will never explicitly point out your errors. That task is left to your judgment, and that is why this document has been written: to allow you to make informed, correct decisions on the basis of the visual display. Your powers of sound perception and judgment are far better than any machine's, and they will be increased through understanding of the relationship between the visual display and your speech. You will find that proper interpretation of the display will automatically instruct you in the ways to further improve your speech!

For the purposes of simplicity of exposition, I will discuss each type of display and group of word-lists separately, and will move from those displays which are more simple to describe to those that are more complex, though this may not be the actual route through the word lists that you will use during your training.

Intonation Displays: Voice Pitch Plotted Against Time

As you recall, you wear two microphones when you are working with the system: a small one near your mouth, where your voice is picked up for the tape-recordings, and a tiny one taped on your throat. The throat microphone cannot "hear" all the details of speech that issue from your mouth, but it is very good at listening to what the vocal cords in your throat are doing. A little reflection will tell you that you have two basic means of producing speech sounds: by making your breath pass through narrow openings, thus producing noise, and by producing vibrations of your vocal cords. These two basic means of sound production are used for consonants and vowels respectively, and of course they occur very often in combination.

When we speak of "tone of voice" or "voice pitch," we cannot refer to sounds produced without vibrations of the vocal cords. Only those sounds (vowels and semivowels) produced while voicing is in progress can be said to have a "tone" or "pitch." In the intonation displays, the two microphones are doing entirely separate things. The mouth microphone is used only for the tape recording, and has nothing to do with the picture displayed. The throat microphone is connected to the display through the computer, and only when your vocal cords are vibrating will a line be drawn.

The displays of intonation are not the simplest in terms of the quantity of speech to be produced, but they are straightforward in explanation since the tone of the voice is directly shown

on the screen. A high-pitched voice produces points that are higher on the screen than a low-pitched voice does. Sections of utterances that are higher in pitch than others will plot as "humps" on the display. The longer the voice is on, the longer the line plotted for the utterance. This display is relevant to rhythm, timing, emphasis, and stress in English speech. It is usually found that words or syllables that are important within sentences are longer and/or have higher pitch than neighboring parts of the utterance. The pitch display makes most, if not all, of these parameters visually explicit for your use. As will be illustrated, the MATCH button moves the subject trace vertically to superimpose upon the teacher's trace. This MATCH facility sometimes operates differently for different word lists, but for intonation, it will always move vertically.

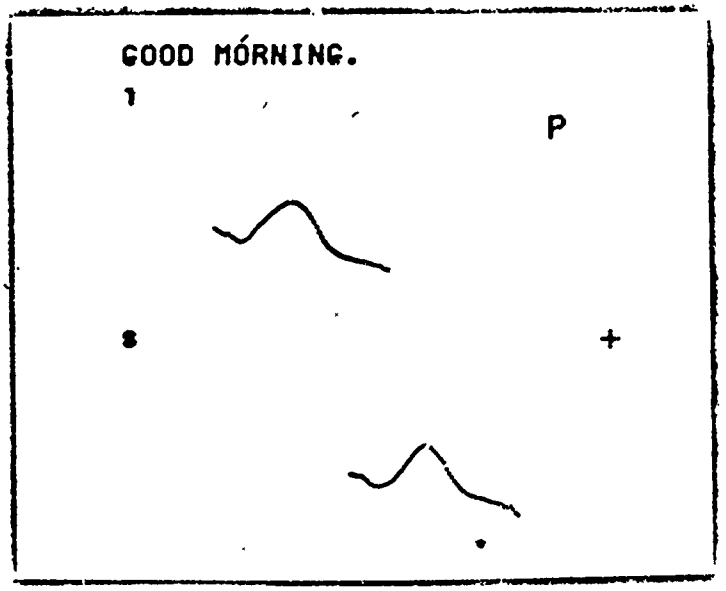


Figure 1a

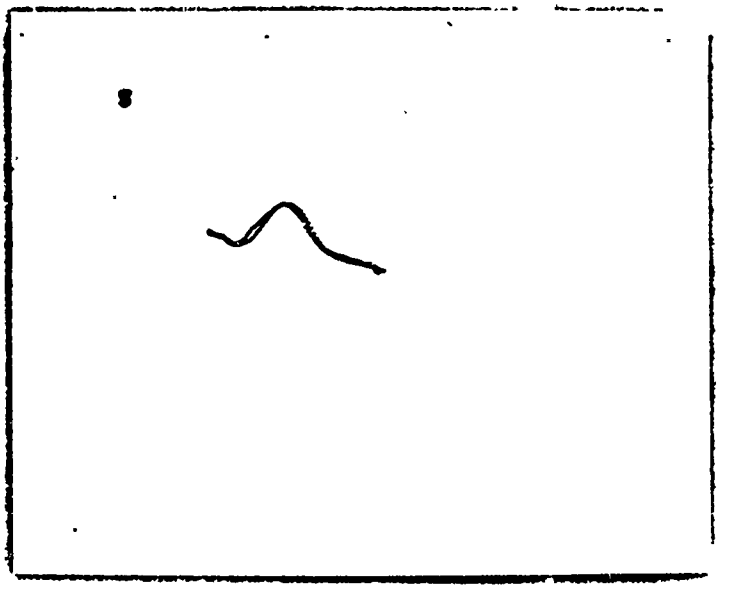


Figure 1b

Figure 1a illustrates an impossible and yet instructive situation. The utterance is "Good mórning". The accent indicates that that syllable is to be stressed in the utterance. It is, of course, difficult to describe and explain the displays in the absence of your being able to hear the sounds that produced them: but let us at least make the effort. The top trace in Figure 1a is the teacher's utterance of the target sentence. The hump in pitch comes somewhere around the middle of the utterance. The bottom trace was produced by the same speaker, but while he acted as a student.

One of the first things to notice about this picture is that the start of the student's trace is somewhat to the right of the teacher's. This means only that the student began his utterance slightly later in time than the teacher did, and is otherwise completely unimportant. As long as the utterance is produced soon enough after the pressing of the STORE button, relative placement within the "time window" is unimportant. Other than that time displacement, the overall duration and shape of the student trace is quite similar to that of the teacher. This is, of course, to be expected since the voice is the same in both instances. The very close similarity between these two utterances is shown by Figure 1b, photographed during the operation of the MATCH button. The student trace was translated upwards and leftwards until the leftmost point of the student was superimposed over or on the teacher's leftmost point. Thus, any time asynchrony between the onsets of both traces has been removed. Any extra length of one function past the ending of another indicates that the two utterances are of unequal durations. Any difference in their shape indicates differing patterns of stress or intonation. The closeness of the two traces here certainly forms a pattern that you might aspire to in your own matches with the teacher, but it certainly is not the only acceptable type of match.

One thing that you must understand about this vertical mode of matching is that the vertical alignment of the traces is only a coincidence between your tone of voice and the teacher's tone of voice. It is the shapes that are important and not their absolute vertical locations during the match. Since the speaker was the same on both sides of the display, it is certainly possible for the absolute voice tones used to match perfectly, but consider Figure 2a. Here again, the same speaker acted as both "student" and "teacher". In fact, the same teacher utterance was used as in Figure 1, but the "student" spoke the utterance in a much higher tone of voice than he had previously. This is not immediately obvious from Figure 2a, since the shapes look roughly similar, though the latter portion of the "student's" trace slopes downward at a slightly greater rate than that of the teacher. The hump in pitch at the accented syllable remains present also. The utterance was produced a bit sooner in time during the store interval, and therefore it lies more to the left than its counterpart in Figure 1.

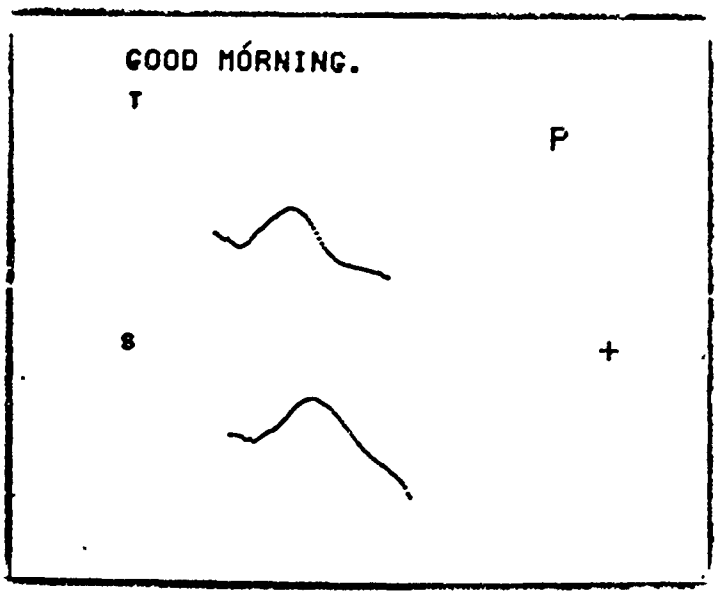


Figure 2a

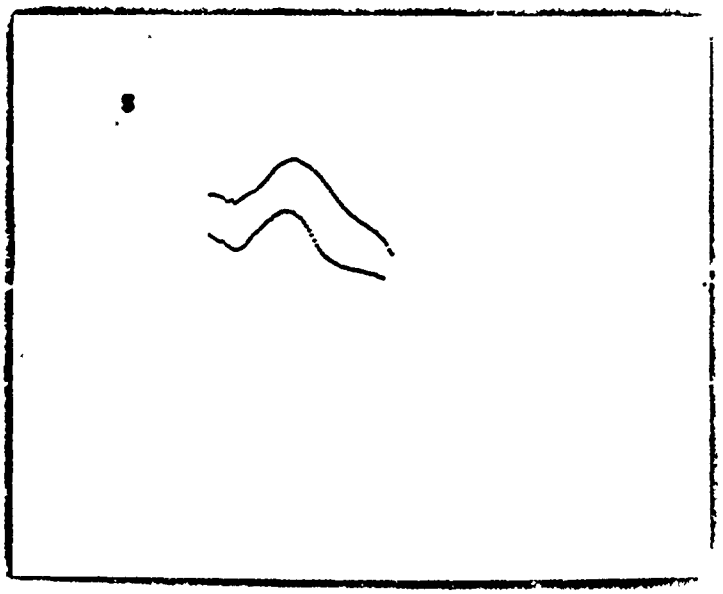


Figure 2b

The dissimilarity in pitch is made immediately obvious in Figure 2b, photographed during the MATCH procedure. The leftmost points of the two traces have been brought into synchronization along the time axis, and the disparity in pitch becomes apparent. The differing shape of the tail end of the student's utterance from the teacher's is apparent here even though there is a large displacement between the two traces. Disregarding for the moment the possible error involved in the trace dissimilarity at the conclusion of the utterances, the rest of the match is indeed a perfect one, even though the two traces do not superimpose exactly. The general principle involved here is this: with vertical MATCH, parallelism is as good as superimposition. Partial parallelism is better than none. It is after all, impossible for a female speaker to match the pitch trace produced by a male teacher. However, it is certainly possible for her trace's contour to match that of the teacher's both in duration and in variability. In conclusion, then: both of the matches shown in Figures 1b and 2b are quite good, despite the lack of direct superimposition in Figure 2b. If the emphasis had fallen on a different syllable, the shape would have been radically different. If one speaker had spoken much slower than another, the traces would not have superimposed so well as far as length was concerned. If both utterances had not been continuously voiced the traces would not have been continuous, curved lines. On all aspects, these matches are acceptable.

Let us take some further examples of work in the intonation word lists and describe some matches that are not so acceptable and tell the reasons why not. Figure 3a shows the speech of two speakers: the teacher being the same speaker as displayed previously, saying the word "airplane". The student is of

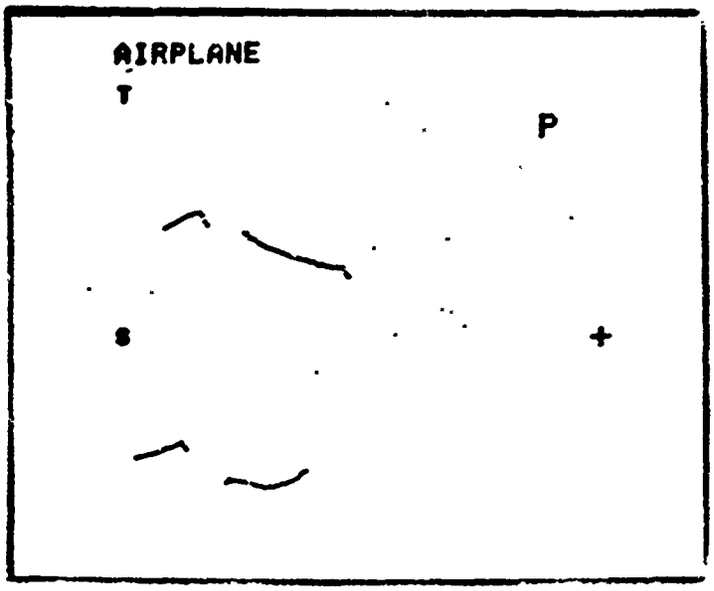


Figure 3a

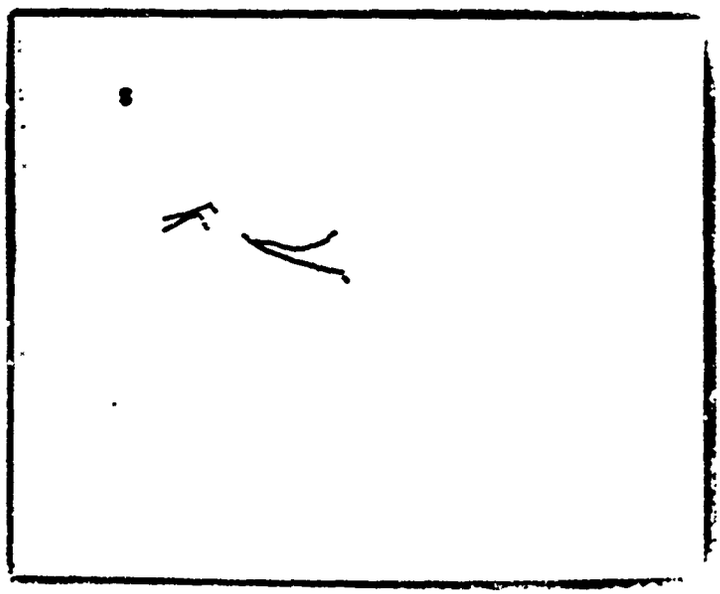


Figure 3b

Spanish-speaking background. His version of the utterance is different in several important respects, the most striking of which is the shape of the contour of the second syllable. Both pitch contours are not continuous because of the / p / in the middle of the word which shuts off the voice for a short period of time before the second syllable begins. This type of break in the voice trace is often useful to you in determining which part of the trace belongs to which section of the word. You can, of course, also use the REPLAY button which will disclose the trace at the same time as the corresponding section of the word is heard over the loud speaker. The student's version of the word appears a little bit shorter than that of the teacher, in addition to the difference in the shape. Figure 3b shows the appearance of the display during the MATCH function. Here, the difference in second syllable shape becomes immediately apparent. The student's first task in this case is to try to speak the word with the same kind of overall pitch contour that the teacher uses. This

involves stretching the utterance out ever so slightly and changing the way the voice moves during the second syllable. If the student can do this, he will mimic the general pattern of the utterance of single words in English. Again, parallelism is all that is required, not superimposition. If the MATCHed traces were just a constant vertical distance away from each other, that would be adequate performance. Overall duration of the entire utterance and internal breaks should also be as consistent as possible.

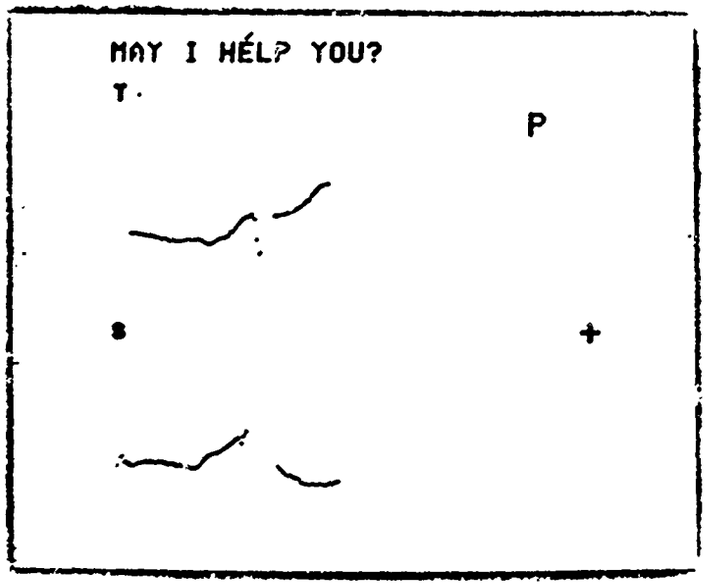


Figure 4

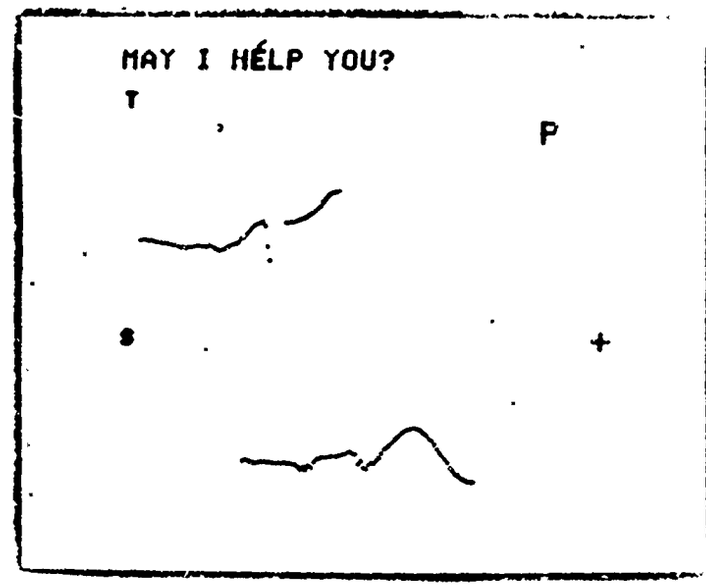


Figure 5

Figures 4 and 5 illustrate some unsuccessful attempts on the part of the non-English native student in mimicking the utterances of the teacher. Here, the sentence is "May I help you?" The break in the pitch trace is caused at the / p / of "help" where the voice stops until the word "you" is uttered. Figure 4

shows a reasonably sized break in the student's pitch trace, but the contour of the word "you" is by no means similar to that of the teacher. Usually, an English question requiring a "yes" or "no" answer will end with a rise in pitch, as shown in the teacher's trace. Since the student's trace does not have this rise, the whole sentence will sound incorrect. The first section of the sentence is not a bad job of pitch matching, but the last word ruins the attempt. (It was not felt necessary to show a photograph of the matched version of this utterance pair.)

It is not always true that the accented word in a question of this sort is the one with the highest-pitched point. The accent on the word "help" is shown as a rise at that portion of the trace, but the question form also causes a voice to rise producing a final syllable even higher than the accented syllable.

Figure 5 illustrates another unsuccessful attempt on the part of the student to match the same teacher utterance. Here, there is no break in the student's pitch trace at the sound of the letter p. Furthermore, the final syllable again does not have the appropriate shape, and in fact it rises and then falls precipitously at the end. This is another obvious failure and would show up as even a worse job in the MATCH operation.

Figures 6a and 6b show an interesting type of error in the utterance of a sentence. As Figure 6b shows, the overall length and overall shape of the utterance is not that different between the two speakers. However, the student's utterance has a break in it and the teacher's does not. The break comes at the word "is". The student produced that word with a pure / s / sound,

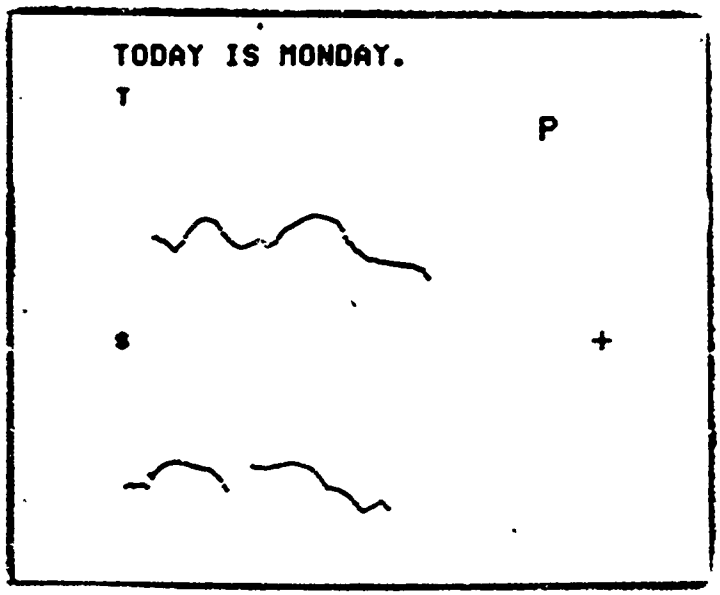


Figure 6a

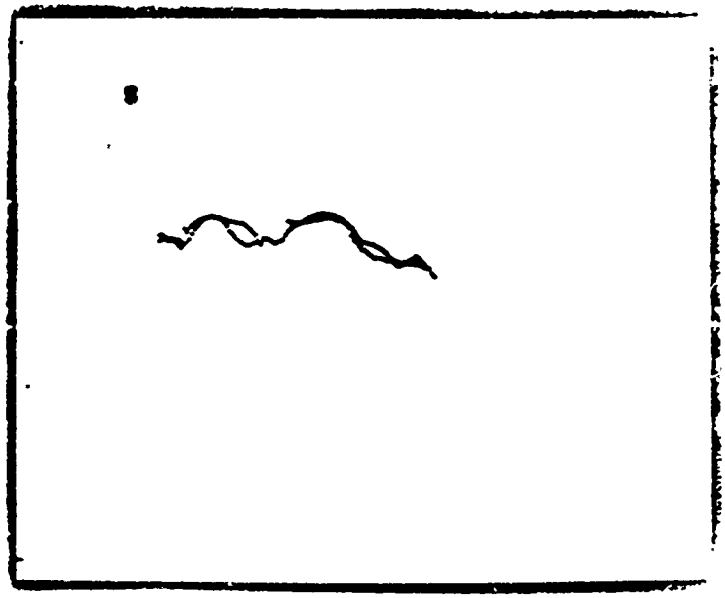


Figure 6b

therefore, making a break in the voicing of the sentence. The teacher's actually had a voiced component (i.e., kept his vocal cords vibrating) throughout, therefore it sounded like the sound "iz". The lesson from this display is, of course, that breaks in your trace should not occur where corresponding breaks do not occur for the teacher. Other than this small discrepancy, the match shown in Figure 6b is rather good. Syllables of the student trace seem to be getting emphasis at approximately the same points in time as those for the teacher.

Vowel Displays: Tongue Position as a Function of Time

When you work with the word lists devoted to vowel training, the system will put very different things on the screen. Instead of the throat microphone picking up vocal-cord vibrations and producing lines during "voiced" speech segments and at heights proportional to the pitch of those segments, the lines plotted here are derived from the mouth microphone. Voice pitch is irrelevant in controlling the height of the lines plotted.

As before, lines will only be plotted during segments of speech when the vocal cords are vibrating (though there may be slight errors occasionally when the system is fooled). Therefore, the horizontal extent of the pictures you draw with your speech will, as before, reflect the duration of the voiced segments of your utterances. The vertical position of the traces is controlled now not by the tone of voice you use, but by the nature of the movement you make with your tongue as you speak the voiced parts of words. In some cases, the trace will move higher on the screen as your tongue moves higher in your mouth; in other cases, for other parts of the word lists, moving your tongue toward the back of your mouth may cause the trace to rise on the screen. A little reflection will show you that one of the major differences between vowels is the position of your tongue as you speak them. Furthermore, if the production of a new vowel involves your finding a new position (or movement) of your tongue that you cannot find using your ears alone, the additional information provided by the display as to where your tongue is and where it should be may be enough for you to learn the new vowel sound!

All the functions of the buttons remain the same. You may still speak, be recorded, and play back the contents of the small loop of tape. The MATCH function works slightly differently, as does the actual display itself in real time. You may speak the training words at one pitch or another, and if the "vowel quality" is wrong, even intonation patterns similar to the teacher's will not eliminate the accented quality of the speech, or improve the display. By "vowel quality" we mean not the pitch at which the vowel is uttered, but rather the nature of the speech sound itself. As you may know by now, English has several vowels that may not be present in your native tongue. Spanish, for example has but five basic vowels, while English has at least twelve. The common

type of error for a Spanish-speaking person learning English is to substitute one of the Spanish vowels for the English vowel. This results in inappropriate vowels in many cases, giving rise to accent. Pairs of English words having different vowels are sometimes not produced correctly by the accented speaker, because he has substituted one vowel for the two which should be distinguished. Furthermore, the problem can be complicated by the nature of certain other vowels in English that are not present in the source language, such as diphthongs, where the speech sound changes during the course of the vowel portion of an utterance.

To help you in making these new sounds and new types of sound distinctions, we are plotting something relevant to the movement of your tongue in your mouth as you speak the vowel portions of pairs of words differing only in the vowel portion. As a little experimentation will show you, changing the position and/or shape of your tongue in your mouth as you speak will change the sound quality of the speech you produce. There are characteristic positions and shapes of the tongue that produce each of the given vowels. An accented vowel is one in which the tongue is not in the right position, or does not pass through the appropriate area by means of the proper route, or is incorrect in duration. All of the above aspects of the vowel are visually displayed for your use in improving your pronunciation. To do this, the display is changed from mirroring your voice pitch to reflecting your tongue position as you speak the vowels.

Consider the examples shown in Figures 7a and 7b: the pair of words is "feed fed". The top speaker, as it always is, is the English native teacher. The bottom speaker is the same student as before, with Spanish as his native language, but with only a moderate accent level. The letters at the upper right,

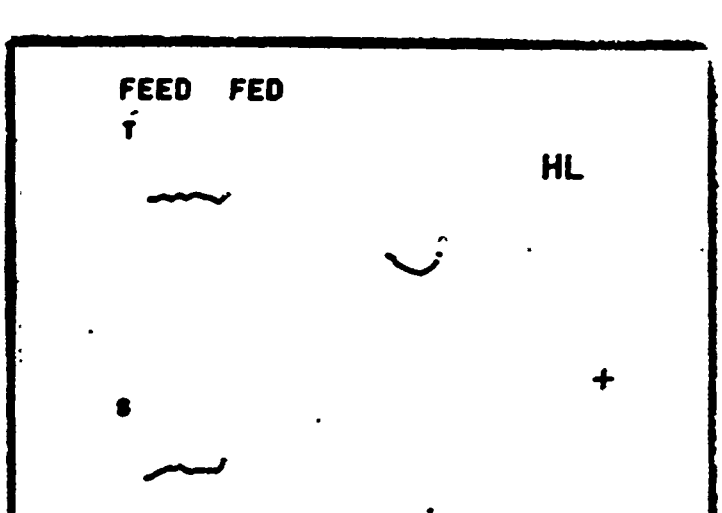


Figure 7a

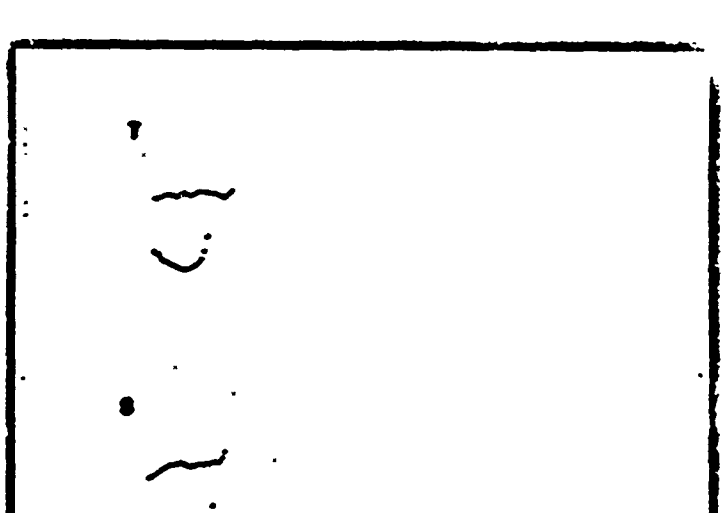


Figure 7b

HL, need not be of too much concern to you, but they do indicate which of the several functions has been plotted on the screen. In this case, what you are seeing is an actual plot of the height of the tongue as a function of time. The higher the tongue in the mouth, the higher the points are on each of the halves of the display. In Figure 7a, then, the sound of the vowel in "feed" is produced with the tongue higher in the mouth than it is in the sound of the vowel in "fed". One sees nothing on the screen corresponding to the consonants of the words that are not produced with the vocal cords vibrating; that is, no lines or points are produced during the / f / or / d / of both words. This will allow you to concentrate on the vowel aspects of the words, although there are some unavoidable effects of the consonants that will be discussed later. All of the word pairs in the vowel lists are composed of two types of words. One member of the pair is always an English word which has one of the five simple vowels that are present in Spanish. Another member of the pair has another one of the English vowels. In all cases, the two words

are genuine English words. There are three basic vowel lists, each having two versions. The word pairs in the straight version are always arranged such that the simple word comes first and the new vowel is second. The scrambled versions alternate that order. Otherwise, the two versions of a given list are identical.

Turning our attention back to Figure 7a, we can see a large difference in tongue height of the two words of the teacher trace, and a similar difference in the traces' height for the student. Now, how good is this performance by the student? It is difficult to assess without being able to listen to the two speakers' performance, but we do have the photograph of the display taken during the MATCH operation. For the vowel lists, the MATCH button does not move the student traces up to coincide with the teachers. This is both unnecessary and confusing. The point at issue here is: How similar is the distinction between the student's two words and the teacher's two words? Since one of the two words has a vowel that is fairly simple for a Spanish speaker to produce, this serves as a fine standard of comparison for the second vowel. Therefore, the MATCH button operates similarly for both speakers, moving the trace of the second word leftward so that its first point coincides with the first point of the first word's trace. As Figure 7b shows, this coincidence does not mean vertical movement, but rather translation of the traces such that the horizontal positions are equal at the beginning of the two utterances. When the traces are in this position, it is simple to inspect the two patterns for discrepancy or similarity. The MATCH shown in Figure 7b is not bad at all, although the distance between the subject's two traces is slightly larger than that for the teacher. Therefore, it might be slightly improved by trying the utterance again with the tongue slightly

higher in the mouth for producing the second word of the pair. The durations of the two words seem to be about the same for the student and the teacher. The "tail" of the student's version of the word "fed" is shaped slightly differently than that of the teacher, but this does not seem to be too large a difference since they both are headed in the same direction at the end of the word; that is, in both cases the tongue seems to be heading upward at the conclusion. Overall, this is a fairly good match. Note further that the difference in silent time between the two words is unimportant. The only important thing is that both members of the pair should be produced within the "time window" available to you as you press the STORE button. The MATCH function will eliminate any time differences between your first and second words, and it will produce a useable double word pattern for you to use in comparing your version with that of the teacher.

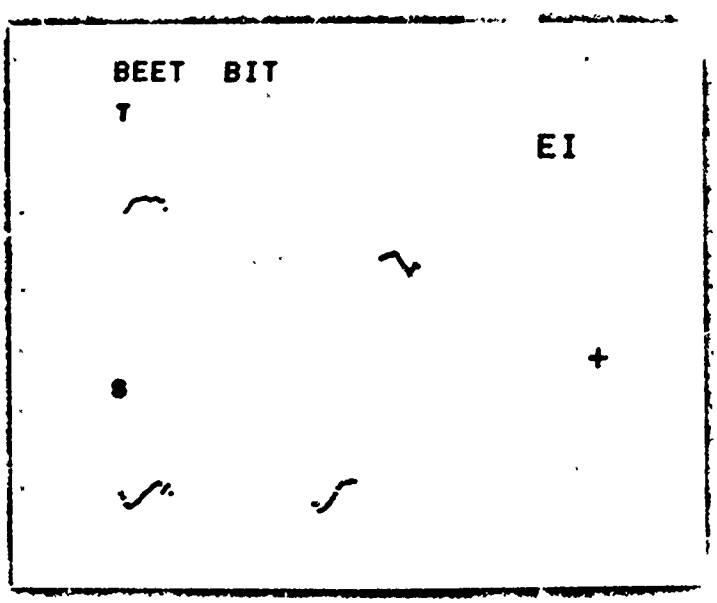


Figure 8a

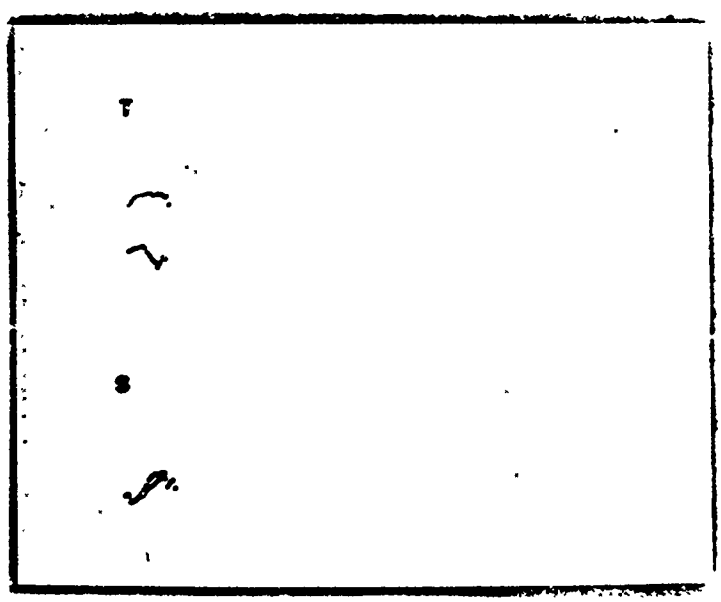


Figure 8b

Let's turn our attention now to another word pair in the vowel word lists. Here, the function displayed is not the HL function but, rather, the EI function. You need not concern yourself too much with the physical meaning of this function, but instead that the "ee" sound as in the Spanish word "sisa" is placed high on the screen, and the "i" sound in the English word "bit" is placed lower on the screen. In general, the higher the tongue, the higher the trace with this function, but at the same time, the more "ee"-like, the higher the trace, and the more "i"-like, the lower the trace. The same two speakers are used in this figure. The teacher's two words are both relatively short in duration as compared to the previous figures. Furthermore, the second word is relatively lower on the screen for the teacher than it is for the student. This discrepancy in vertical position is made clear by Figure 8B, taken during the MATCH operation. If we again make the reasonable assumption that the Spanish-speaking student produced the first word correctly, it then becomes obvious that the second word was not produced with the appropriate difference in tongue position. The tongue should have been lowered during the pronunciation of the second word, and since it was not, the subject's words superimpose on each other in the MATCH display, and the teacher's do not. If <u>you</u> are confronted with a situation like this, you might use the STORE or the FREE mode button to experiment with new sounds that might produce the right kinds of discrepancies. Remember that one of the two sounds will be fairly easy for you to produce. You should know, from your texts or the monitor, the general direction of the tongue from that sound to the new sound. Try to produce that new sound in such a way as to produce the same kind of visual discrepancy as a teacher shows. Try to concentrate on the <u>word</u> pair being trained rather than on the two vowel sounds in isolation, because, if you try the latter, you may be fouled up by the fact that the neighboring consonants change the form of the

teacher's trace to one that you cannot match by producing vowel sounds in isolation. Try to match the trace shapes as well as their vertical position, but do not be confused by the influence of adjacent sounds. This will be further explained next.

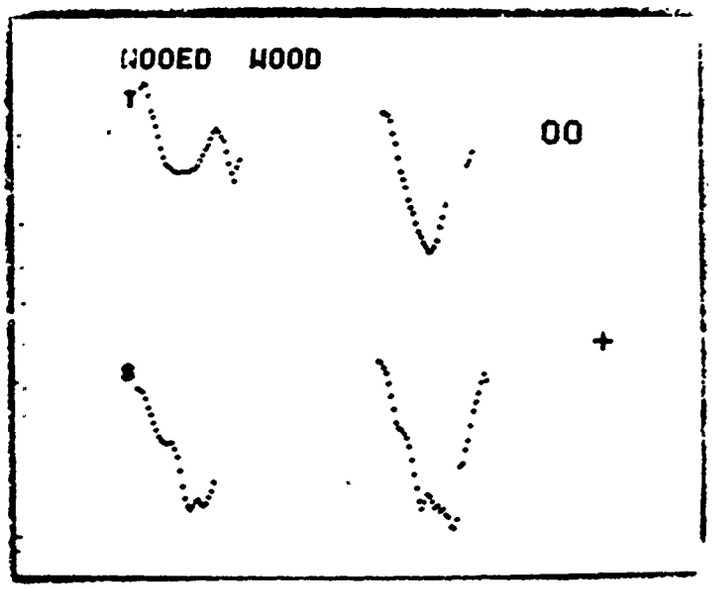


Figure 9a

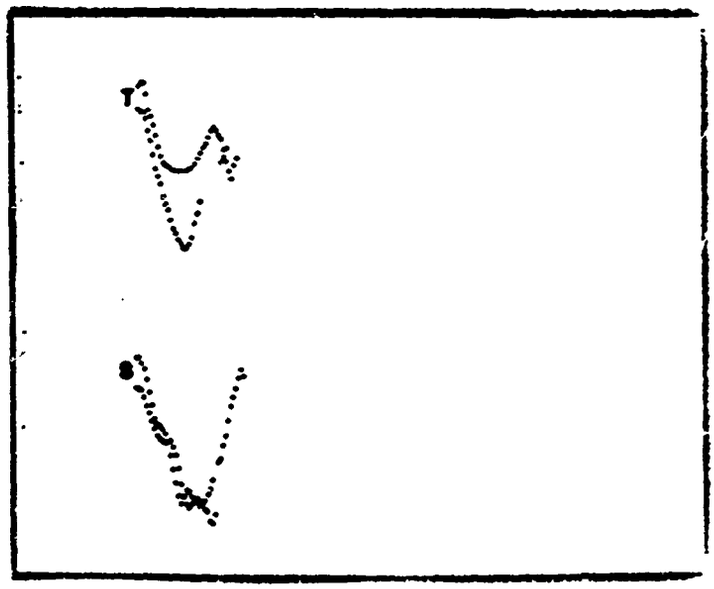


Figure 9b

A good example of the sometimes-confusing influence of neighboring consonants is shown in Figure 9. The first member of the vowel discrimination pair in this figure is the vowel "oo," as pronounced in the Spanish word "su." The second vowel is the shorter "u" vowel used in English. It is produced with the tongue slightly lower in the mouth, and it is usually of shorter duration. The function shown on the screen is labeled "OO" and is specifically designed for this type of vowel pair. The closer the trace is to the original vowel as in the Spanish, the higher the trace becomes. The more toward the English short "u" the

sound becomes, the lower the trace. At first glance, the match produced by the student for this word pair came quite close to that of the teacher; but this is deceptive. It would be especially valuable at this point to be able to let you hear the sounds that produced these two displays, but of course that is not possible. Notice that the leading edges of all four words slope sharply downward. This is caused by the fact that the W in both words is what is called a semi-vowel. It is impossible for the computer to discriminate between a semi-vowel and a vowel, because both of them involve production of similar types of sounds. However, through the use of both the visual display and close attention to the audio REPLAY, you may discriminate between the parts of the display that were produced during the semi-vowel from those parts produced during the vowel itself. You will note in Figure 8B that the central parts of the teacher's two traces are separated from one another, while the central parts of the student's two traces superimpose rather closely. The reason for this is that the student's production was in error. There was no difference between the two vowels in the student's version of the two words, while the teacher's production did show the required difference. The point of the matter is, therefore, to pay the most attention to that section of the display that was actually produced during the vowel itself. The easiest way to do this is to watch the display carefully during the operation of the REPLAY button. Watch carefully, and listen to those sections of the picture and sound produced during the full-blown sections of the vowel; disregard those that happened to be produced during semi-vowels or even, in some cases, during those few kinds of consonants that may produce points on the screen. The latter type of points are also in evidence in Figures 9A and 9B. They can be detected and disregarded by the same means--inspection of the display while the REPLAY procedure is going on.

No further specific comments need to be made about the remainder of the contents of the first or the second vowel lists. The word pairs in those lists are all fairly similar, in that the vowels are short to moderate in length and that when the tongue reaches a given region for a vowel, it tends to remain there for the duration of that vowel. That is, the words in those lists are more "tense" than "diphthongized." The functions displayed on the screen for each of these vowel pairs have been selected to maximize the difference in display position for the two members of each pair. In the cases where the HL or FB functions are displayed, you may think of these most easily as the high-low, or front-back, positioning of your tongue. If the FB function is selected, for example, a high position of dots on the screen is produced by a front position of the tongue in the mouth, vice versa for the back position of the tongue in the low position on the screen. The other functions have a more complex relationship between tongue position and position on the screen. All that need be said for the display of these functions is this: You can produce at least one of the vowels in the word pair without much difficulty. In the straight version of each of the lists, the member that you can produce easily is the first one. Take the trace that you can produce (making sure that its duration and approximate shape seem more or less similar to the teacher's), and use that as a starting point against which to compare your utterance of the second word. Try as many experiments as you wish to produce the same kind of discrepency that the teacher's trace produces. Refer to your textbooks if you need to find out approximately where the tongue should be for this particular vowel, and move in that direction first. After you are able to make sounds that produce the appropriate vertical visual discrepancies, concentrate next on the shape and contour of the trace. Through all of this, make sure that you are attending only to those sections of the display produced during the vowel section, and not to those sections produced during semi-vowels or consonants, as described above.

As you work with these and the other lists, make sure that you can easily and consistently produce the correct discrepancies between your word-pair traces. It is not enough that you managed to do it right once. Can you repeat the performance at will?

Here is another very important thing to remember. Whenever you are using the system, do not depend entirely on the pictures. Remember your ears! Your pronunciation of the words is, ideally, supposed to sound identical to the teacher's, and the pictures are only to help you in learning the correct ways to pronounce the aspect of speech being plotted on the display. If the display shows a good match, that does not mean that all of your pronunciation is perfect. Listen to the whole thing and try to get every part of it to sound like the teacher.

Next, we will discuss the appearance of displays for the third vowel list, the one having to do with diphthongs and diphthongized vowels. It is the nature of these vowels that the tongue moves from one place to another during their production. This may be a very unfamiliar feature of the English language for you, and it may cause some difficulty. However, the teaching machine allows you to see the nature and time course of the gesture you should make with your tongue, and can help you in the production of these strange sounds. As before, one member of each of the word pairs in the third vowel list is a vowel that you will have no problem pronouncing, and the second is one which is diphthongized. Furthermore, one of the components of the diphthongized vowel is always present in the first member of the vowel pair. That means that if you press the MATCH button one section of diphthongized vowel will superimpose on, point towards, or be close to the area of the trace produced by the first member of the word pair. In most cases, the two non-diphthongized vowels from which the diphthong is comprised are used sequentially in the word list to allow you to work on both sides, both ends of the dipthongized vowel.

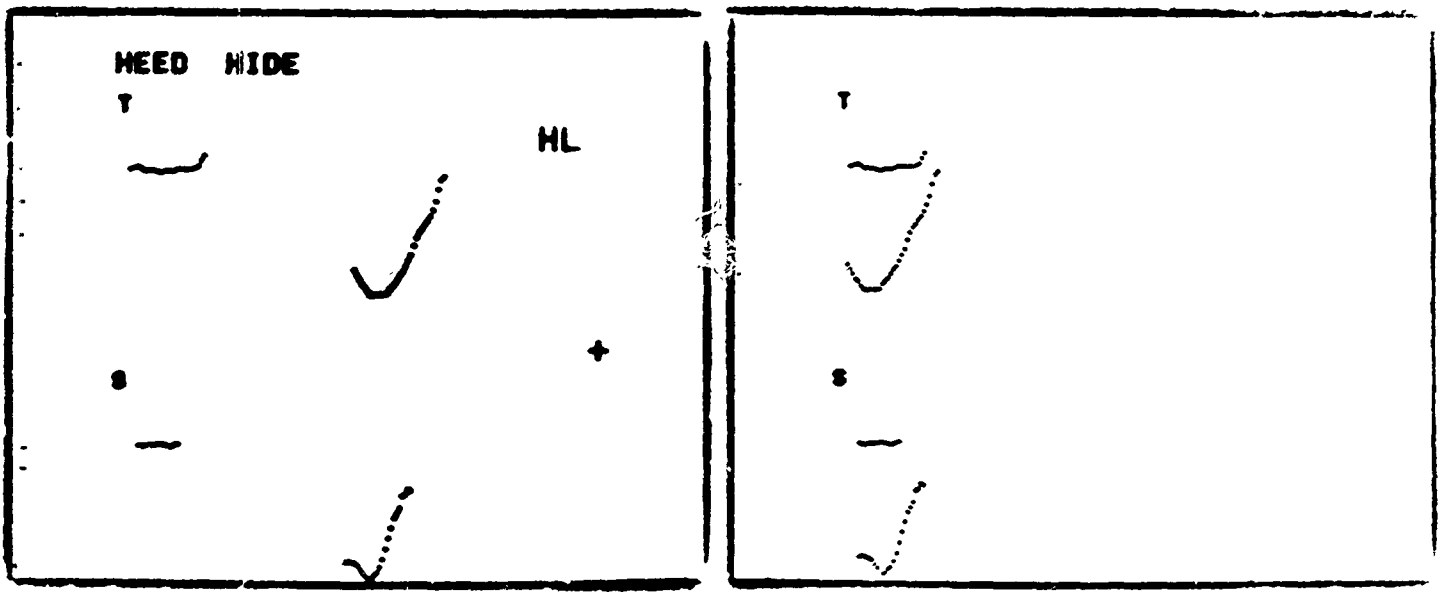


Figure 10a

Figure 10b

Figure 10, again produced by the same two speakers as before, shows a moderately successful attempt at the production of a diphthongized vowel. The high-low function is used, and the general shape of the gesture for the diphthongized vowel is the same for the student as for the teacher. However, there are some points that should be brought to the student's attention. First, the duration of the first word on the student's pair is shorter than that of the teacher. This should be corrected. Second, as may have been made more clear by Figure 10b, even though the gesture shape of the diphthong is similar to that of the teacher, the trace has not reached the immediate vicinity of the trace of the first word as it does in the teacher's trace. The student's version of the word "hide" must therefore have not had enough of the "ee" vowel in it, for if it had, the trace would have achieved the height shown for the first word. If the error in duration and the error in vertical extent were to be corrected, the utterance would probably have sounded far better.

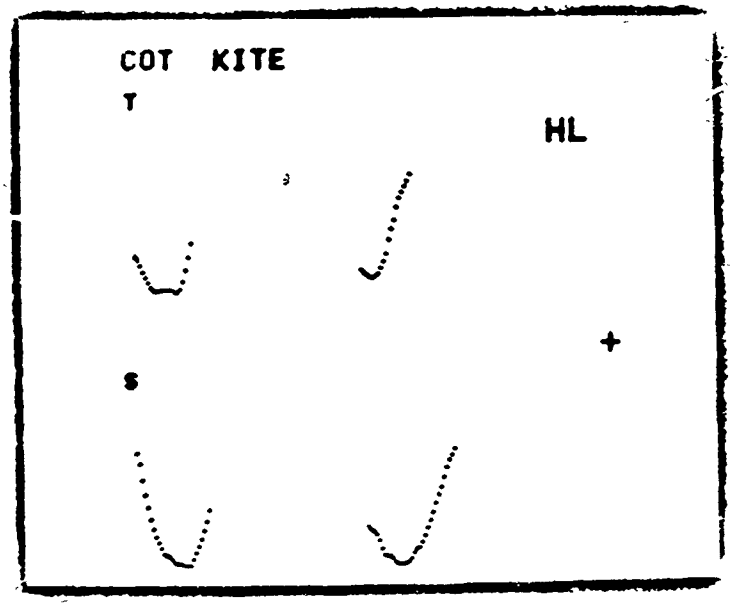


Figure 11a

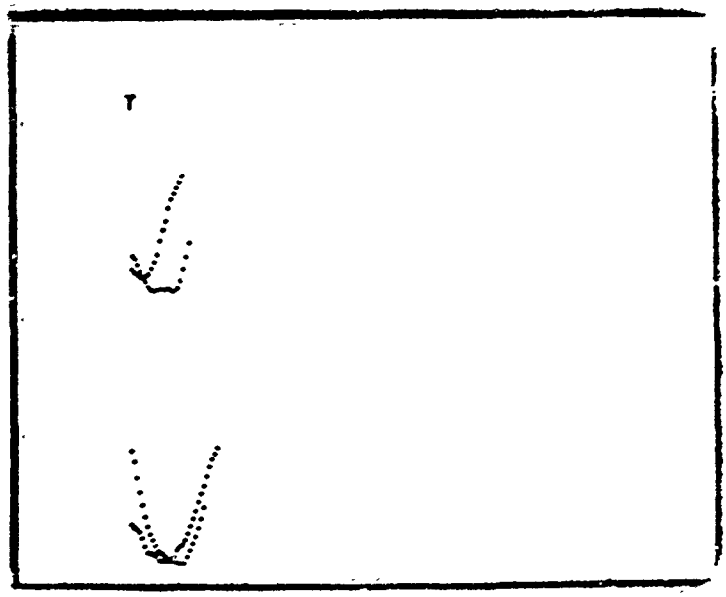


Figure 11b

Figure 11 shows a more successful attempt at matching by the student. This word pair also uses the high-low display function, and with that function one can see that the duration of the student's utterances more closely approximate those of the teacher. For the word pair, the first member of the pair is the vowel that starts the dipthongized vowel of the second word. Consequently, it is necessary that the beginning part of the second vowel occupy the same vertical area of the display as the entire section of the vowel of the first word. As figure 11b shows, this condition is met. The rather large initial curving segments of the student's first word may have been caused by the onset of the consonants, and should be neglected. The one questionable aspect of this performance is that the student's second word may have, in fact, been slightly too long. This is made most obvious by inspecting Figure 11a, and leads to the point that a good deal of matching information can be achieved by inspecting these displays before the MATCH button is pressed, in addition to the information it can give one while in operation.

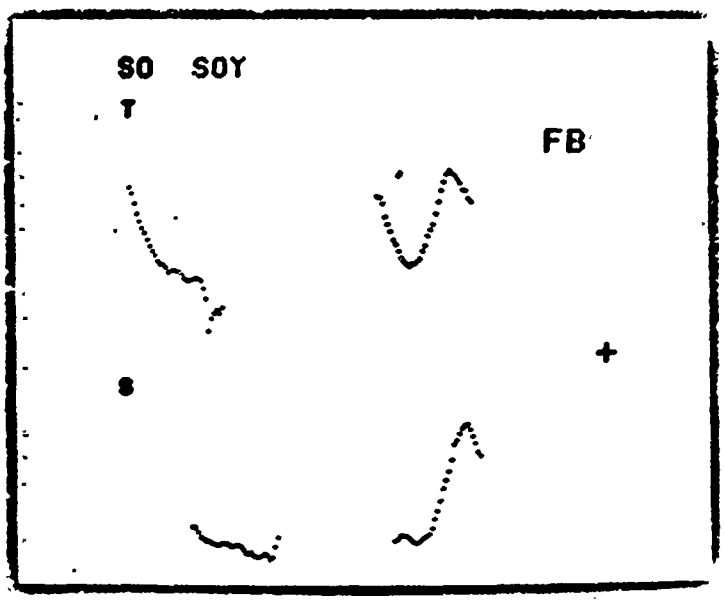


Figure 12a

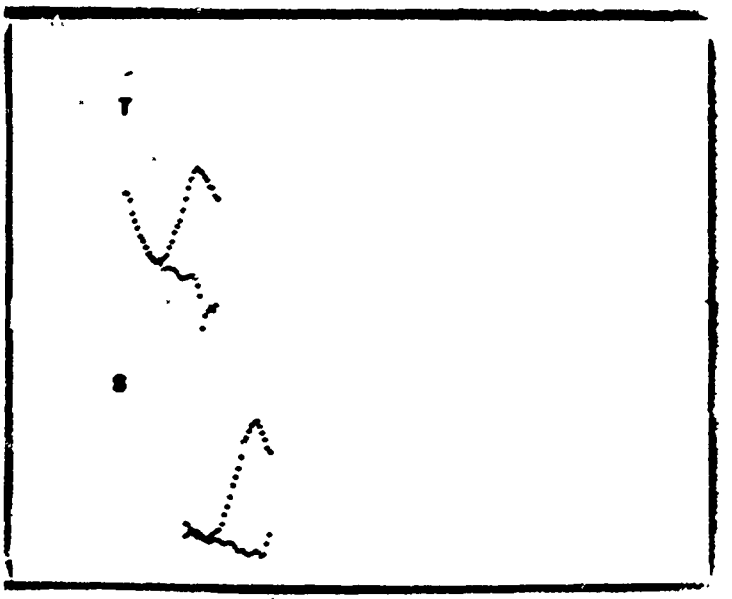


Figure 12b

Figure 12 illustrates a good, but probably confusing, match. Again, the same speakers appear. The student has matched the durations of his traces fairly well with those of the teacher. However, the teacher's two traces seem to have a fast falling initial section coming before the rest of the traces, which are similar to those done by the student. Other than that, the first section of the diphthong of both student and teacher is in the same vicinity as that for the entire main vowel section of the first word, and that is as it should be. Since the front-back function was used, that means that the O in "so" is further back than the conclusion of the diphthong in the word "soy". While watching this display and listening to the replay of the four utterances that make it up, it was apparent that the leading edges of the teacher's words were produced between the transition from the S to the vowel, and that this kind of trace was somehow not produced from the student's speech. The lack of that effect does not detract from the goodness of the student's performance in this case.

Consonant Displays: Pitch and Loudness as Joint Functions of Time

For all of your work on consonants, the display will act in the same way: it will combine immediate feedback about your pitch with a new function, the loudness of your speech. You are already familiar with the notion of a pitch display from the previous description of the intonation lists. There are three consonant lists, each of which exists in both a straight and a scrambled version. The nature of the difficulties in the first list can be roughly described as not making the proper type and duration of noise with your vocal apparatus before the vocal cords begin to move. You can make noise with your vocal apparatus in two ways: by producing voiced speech, such as vowels, and by producing unvoiced speech sounds, such as the sounds in the letters S, F, and the shorter, noise-like sounds of the letters T, P, and so on. In the consonant lists, we will concentrate our attention on the proper manner of production of these kinds of unvoiced speech sounds. The display that we use to do this is more complicated, and on occasion the distinctions which you will be asked to be sensitive to are rather subtle. Therefore, be sure you understand the following comments, and ask the Monitor for any further clarification you feel you need.

Let us inspect Figure 13, a picture of the display after the same student had imitated the same teacher in the production of the word pair "deem-team". The letters PL signify that the display shows both the pitch and the loudness of the speech at the same time. If a speaker is quiet, nothing appears on the screen for that instant of time; if he is making a noise in the absence of the movement of his vocal cords, a single line near the bottom of the section of the screen devoted to that

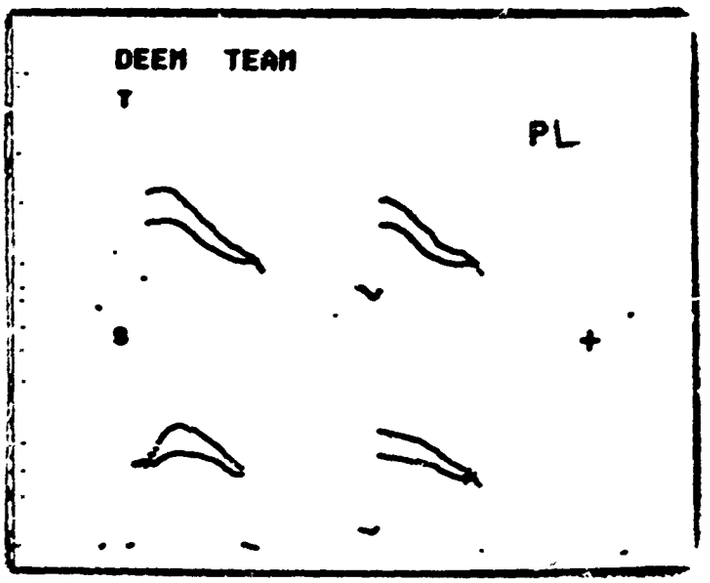


Figure 13

speaker is plotted. If he is making a speech sound that is produced with the action of the vocal cords, this produces both a pitch and loudness display, and is indicated by a double line. As before, the display is divided into two halves. The bottom half still belongs to the student, and is changed not only because of the new display, but by the addition of two small dots at either extremity of the bottom of the display. The same holds for the teacher's half. The small dots are added to help you tell the difference between the pitch and the loudness functions. As you watch the display, draw an imaginary line between those two dots. That line indicates silence. Try the experiment of making a continuous "S" sound while the system is listening to you, and you will find that you will draw a line starting the instant that your S sound begins, whose height above that imaginary line is determined by how loud you speak the "S" sound. Continuing the experiment, changing your speech to a "Z" sound will immediately cause the display to jump upward. Two lines will appear. The bottom line is your pitch, and it will change

position according to what pitch you are using in producing the "Z" sound. The top line is the loudness of that part of your speech, and it will be plotted at a distance above the pitch line proportional to how loud the speech is. If you "wobble" the pitch of your Z while maintaining the same loudness, you will draw two wavy lines that are roughly parallel to one another. If your Z sound changes in loudness but not in pitch, than the bottom line will remain roughly straight while the loudness line, plotted above it, changes its distance from the bottom line. The quantity plotted for the "noisiness" of both the S and the Z sounds is the same: we have called it "loudness" above, but you might also think of it simply as the amount of noise you make while you speak, independent of whether it is a voiced or an unvoiced sound, or if it is voiced, with what pitch it is uttered. The louder the noise picked up by the microphone, the higher will the loudness points be plotted. There is, of course, a distinction between the loudness of voiced and unvoiced sounds. That is the major distinction that this display allows you to see. If the vocal cords are not in operation, then loudness produces points at a variable distance away from the invisible lower base line connecting the two extreme dots. If a noise is picked up while the vocal cords are in operation, then the amount of loudness is shown as the distance away from the pitch trace.

It is well to remember that the two lines come from two different portions of your body: the pitch line is produced only when your throat is vibrating such that the small microphone taped onto the throat is activated. When this happens, the pitch line is plotted at a height proportional to the pitch that you are using. The loudness is obtained from the voice microphone itself. It therefore is sensitive to both noises produced without the vocal cords and those noises produced with their

help. As a general rule, noises produced without the vocal cords are fairly short in duration, and arise mostly from short consonants. When these consonants occur at the beginning of words, they often produce a small amount of noise (shown by the loudness function alone) before the vocal cords begin to operate in the following vowel. It is the presence and the length of this introductory burst of noise that concerns us especially in the first consonant list, to which we now turn our attention.

Let us inspect the teacher's half of Figure 13. The first word in this word pair begins with the letter D. The teacher has produced this in the following way: there was a small amount of noise before the vocal cords began, but so small that only one lone point of the loudness function was plotted near the invisible base line before the vocal cords began operation. As soon as the voice began to operate, the pitch gradually fell as did the loudness, producing the falling, tapered form shown for the teacher's first word. The second word starts with the letter T. There is a good deal more noise before the vocal cords begin to operate. This noise is produced by the rushing of air through the mouth between the time that the T is begun and the onset of the vibration of the vocal cords. Proper English pronunciation requires the presence of a fair amount of this noise before the beginning of the vocal cord activity. A proper English distinction between the letter D and the letter T requires that the letter D not have too much of this vocal tract noise before the vocal cords begin, and that the letter T have a good deal more. The feature of major importance, then, in the teacher's utterance of this word pair is that the loudness plotted for the first word before the vocal cords begin vibrating must be much smaller in extent than that plotted for the second word. Let us see how well the student has done in matching this feature. At first glance, the student's performance might seem unacceptable, but it is in

reality quite adequate. The amount of noise present before the onset of the pitch in the second word is a close approximation to that of the teacher. The difference in the appearance of the first word is caused by something known as "pre-voicing". While the subject's mouth was still closed, his vocal cords began to move and only a short time later did the mouth open. This is sometimes referred to as "swallowing" the consonant. As long as there is little or no initial loudness plotted before the voice begins, whether the utterance is pre-voiced or no, an acceptable version of the consonant "D" will be produced. English makes no distinction between no pre-voicing and pre-voicing of this type of consonant. Other examples of pre-voicing will be shown below so that you understand the idea. The easiest way to recognize its presence is that the pitch and loudness lines are very close to one another for a short period of time, and then the loudness line diverges smoothly but quickly from the pitch line. At the point of divergence, the pitch line may also change its shape or direction. To finish the discussion of Figure 13: the subject's match is adequate, since the amount of noise in the second word plotted before the voice begins is much larger than is the amount of noise for the first word.

For all the consonant lists, the MATCH function operates in such a way as to move the two pairs of words vertically, as it did for the intonation displays; however, since you are dealing with word pairs here, the vertical match treats each word separately. That is, your second word is translated upwards so as to superimpose on the second teacher word, and the same for the first word in the word pair. This was not photographed for the handout because of technical difficulties, but it is of some use for you in comparing the length and presence of the amount of noise before the vocal cords begin activity.

For all of the consonant lists, you will see the pitch-loudness display. The major reason for the presence of the pitch line is to make clear the distinction between voiced and voiceless speech noises. The object of most of your attention should be the proper production of those noises. If, however, you wish to concentrate also on the pitch contours, and on the syllable duration, that information is also available for your use. If you are able to pay attention to these aspects, try not to worry too much about exact matches of pitch. Remember that it may not be possible for you to exactly superimpose your pitch upon that of the teacher. The overall trend of the pitch in a syllable is of some interest, and you should also take some care that your overall word durations are roughly similar to that of the teacher. You will not find any feedback here on the proper vowel quality. That work was done in the previous vowel lists. Again, remember your ears and concentrate on the entire sound, not just the consonants.

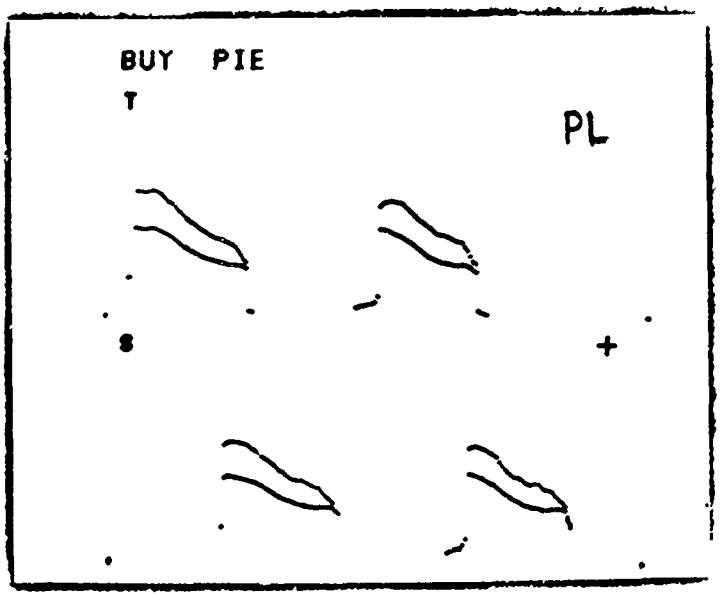


Figure 14a

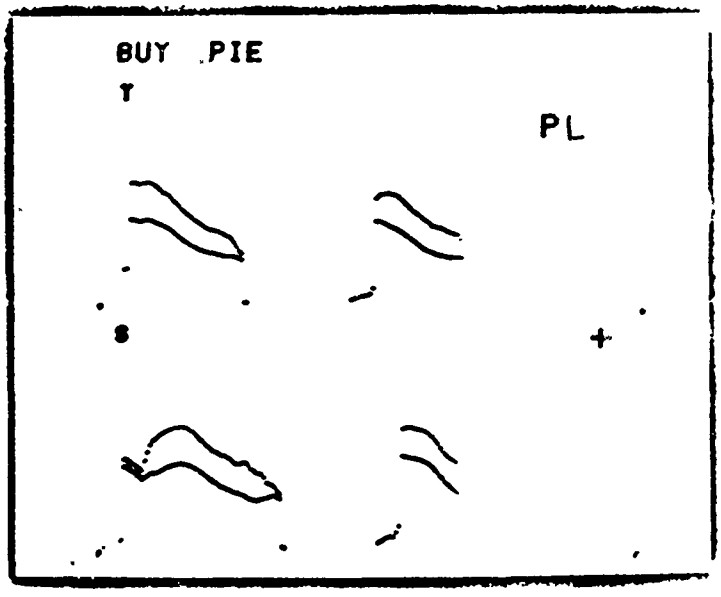


Figure 14b

The next figures, drawn from the next section of the Consonant 1A List, illustrate two good matches between the student's performance and the teacher's performance. The teacher does not pre-voice his B in the first word, and the second word has a large amount of noise before the onset of the pitch. The student's display for the first word has many fewer loudness points before the onset of the voice than does the second word, and the amounts shown are very similar to those of the teacher's, and therfore the match is quite good. Incidentally, the pitch contours and durations are also quite comparable. Figure 14b is a good example of an acceptable match with pre-voicing on the B. All the signs for pre-voicing are present here; the small value of loudness plotted over the pitch contour, the fast rise in loudness at the opening of the mouth, and the shift in the position of the pitch contour at that point in time. Even before the pre-voicing begins, there was one loudness point plotted. The second student word has the right amount of noise before the onset of the vocal cord activity. Therefore, since the distinction in the number of points before the pitch begins is the same for both student and teacher, irrespective of the pre-voicing, the match between these two word pairs is quite good.

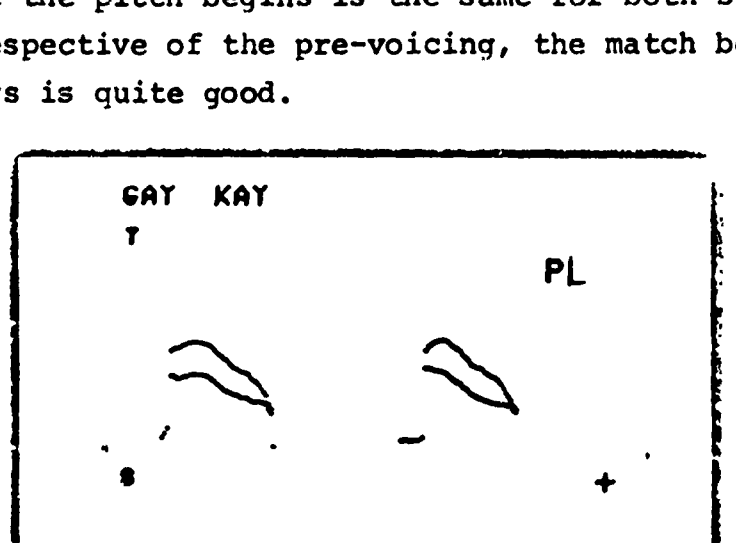


Figure 15

The final example from the first consonant word list is also a successful attempt on the part of the student. First, for the teacher, we see a large amount of noise for the second word, far more than for the first word. There will be occasions where the supposedly "non-noisy" consonant does produce some activity before the onset of the voice, even for the teacher; but in those cases it will inevitably be true that more noise will be present for the "noisy" member of the consonant pair. The student's version of the word "gay" shows some pre-voicing, and his second word shows a small amount of noise before the onset of the voice. The latter quantity of noise might indeed be improved, but at least the distinction is in the proper direction as compared with that of the teacher. Even when such a distinction is possible for you, you should always attempt to produce the same amount of noise activity as does the teacher, although the exact shape of that noise is unimportant, only its duration. Make sure that the consonant that should get most noise gets more than that of the other consonants. This is the central point for this entire word list, both scrambled and unscrambled versions.

We turn now to the second consonant list. Here, the distinctions of interest may not always be at the beginning of the word, because the word pairs may differ in a consonant that occurs in the beginning, in the middle, or at the end of each of the two words. The visual displays produced by these consonants sometimes differ only slightly, and you must know exactly what to look for in terms of the relationship between the loudness and the voice pitch traces. This is especially true when the consonant of interest occurs in the middle of the word. The distinctions between one consonant and the other are more complex in the second and third consonant lists than they were in the first, but the things you learned in that list about pre-voicing and the noise you produce before the vowel will still be of aid to you in the final two consonant word lists.

In the second consonant word list, the major difference between the members of each word pair is that one of the members of the pair has the consonant that somehow momentarily stops the action of the voice before the word continues, and the other is more continuous through the entire word. This small interruption of the word, produced by in some way blocking off the path from the throat to the lips, has its effect on the display produced. If the interruption is produced during a period where the vocal cords are vibrating, then the presence of that interruption often produces a small discontinuity in the pitch function. If the interruption is produced at a time where the vocal cords are just starting or ceasing activity, the perturbation may still be there, but may be harder to see. At that time, the momentary blockage produces its strongest effect on the loudness line. Since, for the duration of the constriction in the voice, the amount of noise coming out of the mouth is minimized, the loudness function will also have a "notch" cut out of it. If the constriction comes at the beginning of the word, then the loudness function will grow more rapidly after the constriction than if

there had been no constriction present, as in the other member of the word pair. Similarly, if the constriction comes at the conclusion of the word, the loudness function will decrease more rapidly because of it than if it had not been present. So, in the following examples, let us pay particular attention to exactly what the loudness and pitch traces are doing at the instant of the constriction caused by the consonant in question. In general, the two words in every word pair are approximately the same length. In working with this and the next word list, it would probably be a help to you if you tried to produce both members of the pair with approximately the same duration, and, as far as possible, with the same pitch contour for both words. You will notice that the teacher attempts and most often succeeds in doing this.

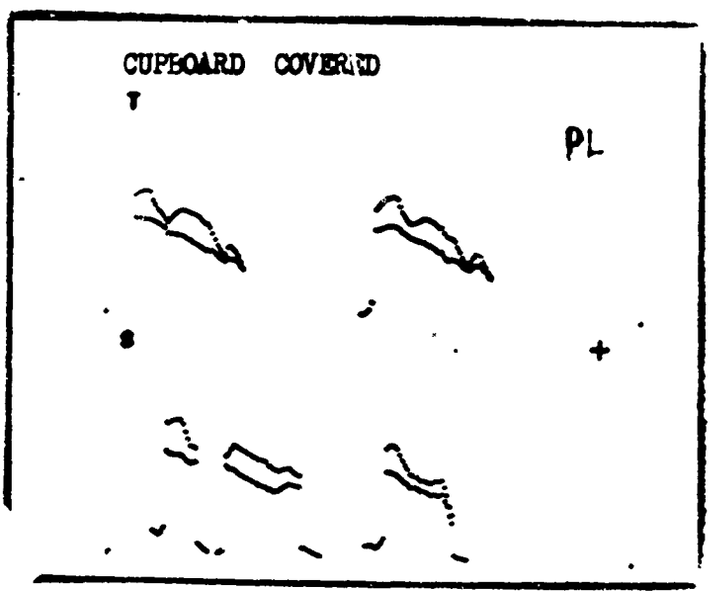


Figure 16

Let us concentrate first on the teacher's version of the above word pair. The points indicated by arrows labeled 1 and 2 are the sections of these two words produced by the consonant discrimination to be trained by this word pair. The constriction produced at arrow #1 was produced of course by the consonant B. Note the sharp discontinuity in the pitch trace, and the fact that the loudness trace is affected before the pitch discontinuity occurs. Contrast that section of the first word with the section pointed out by arrow #2. Here, there is no discontinuity in the pitch trace, and the loudness discontinuity, while present, is much more smooth in its onset and offset. This is because the consonant "V" is produced with a smaller, less severe constriction of the vocal cavity than is the consonant "B." All other sections of the display are not relevant to this consonant discrimination. Let us look now at how well the student did in imitating this utterance. The answer is: not well, but not terribly. The break in the pitch trace indicated by arrow #3 shows that in the student version of the work cupboard there was a total stoppage of the voice between the two syllables of the word. The second word, covered, indicated by arrow #4, is of much shorter duration, and there is no apparent constriction of the loudness trace caused by the consonant V. Therefore, the distinction, while possibly present, has been overdrawn.

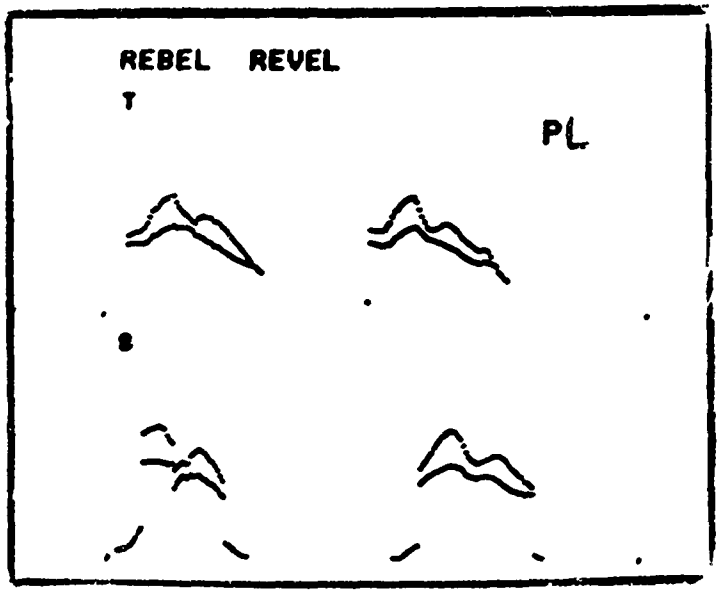


Figure 17

The next figure illustrates a successful match by the student for a word pair having the consonant discrimination in the middle of the word. First, in the teacher's trace, note the sharp discontinuity in the loudness function caused at the consonant B in the word "rebel," and note its absence in the word "revel." The student's trace indicates a loudness and pitch discontinuity for the first word, and smaller discontinuities for the second. Again, since other aspects of the word are irrelevant, the match is a fine one. Note also that the student's duration of voicing is approximately the same as that of the teacher. The initial and final sections of the student's word, where loudness and not voicing is indicated, are not of any concern at this point.

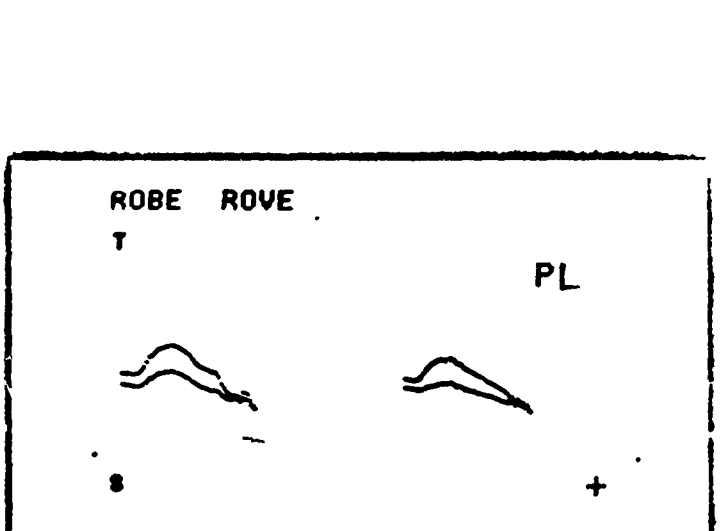


Figure 18

The next example involves a word pair where the consonant distinction comes at the end. Note first in the teacher's trace the B at the end of the word causes the usual change in both pitch and loudness traces. The "V" at the end of the second word produces a slow decline in loudness and no sudden change in pitch. The student trace also shows "bumps" in both functions for the first word and a smooth decline at the end of the second word. The fact that the subject's second word has a small bit of loudness in the abscence of pitch at its conclusion should not be of too much concern. The major thing to note when the distinction occurs at the end of the word is that the word containing the "B" should have the discontinuities in the functions and the word containing the "V" should be without such discontinuities. The durations of the subject's two words seem to be more or less correct, although the contour of the first word is not quite the same as that of the teacher's. The student's second word seems to be a fine match to that of the teacher's, with the above-mentioned exception of the loudness without pitch at the end.

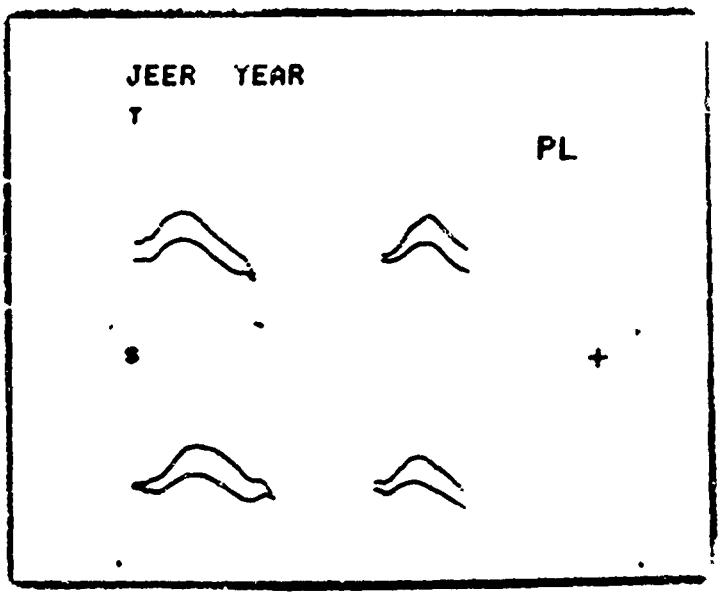


Figure 19a

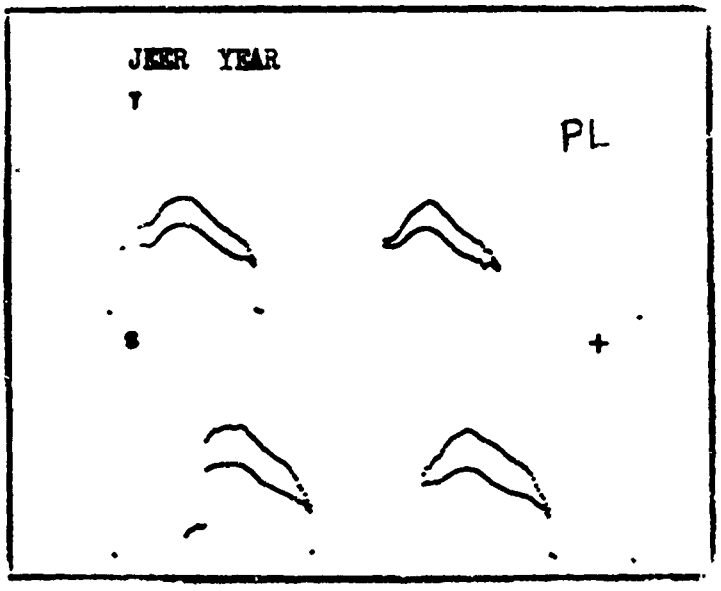


Figure 19b

The middle section of consonant list #2 involves consonant discriminations that lie at the beginning of the word pairs. The consonant "J" is formed by constricting the vocal cavity and opening it while making some noise with the breath through the opening as it appears. At around the same time, the vocal cords are activated. The consonant "Y" is produced with the mouth already open, and the vocal cords are activated and the loudness of the sound builds up rather gradually. An unsuccessful student match is shown in Figure 19A. First, let us inspect the teacher's trace for this word pair. Note that the teacher's version of the word "jeer" starts with the simultaneous presence of loudness and pitch. The loudness does not change too much as the pitch moves from the J to the following vowel. In the second, the loudness starts out low and gradually rises as the utterance changes from the introductory "Y" to the following "E" vowel. As you can see, the consonant "Y" is in reality a semi-vowel. The major distinction, then, is that the consonant "J" is produced with a large amount of noise at the beginning of it, while the consonant "Y" is characterized by a gradual increase in the loudness as the utterance changes into the following vowel. The student's attempt at this word pair is not successful for the following reasons: His first word starts with a gradual increase in the loudness function, whereas it should have had a more rapid onset. The second student word is not bad at all, but the distinction between the two words was not properly made. Figure 19B is not intended to be a good match, but it does indicate an important point. The same teacher recording was used for Figure 19B as 19A, but the teacher also produced the two utterances at the bottom half of the display. The two words spoken were two different, but acceptable, variations on proper pronunciation of the word "jeer." The first version shows noise before the vocal cords are activated. That is OK for the consonant "J," but of course would not be acceptable for the semi-vowel "Y." The second utterance

on the student display was produced with the noise simultaneous with the onset of the voice. Note the fact that the loudness does not increase in the smooth way that the word "year" produces in the top half of the display. In general, then: the first member of these distinctions should produce a rapid, discontinuous onset of both pitch and loudness. The second member should produce gradual onset.

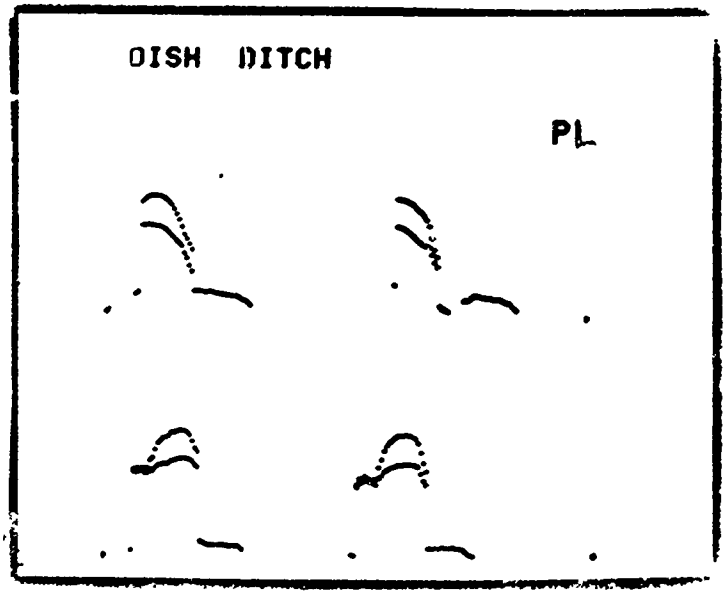


Figure 20a

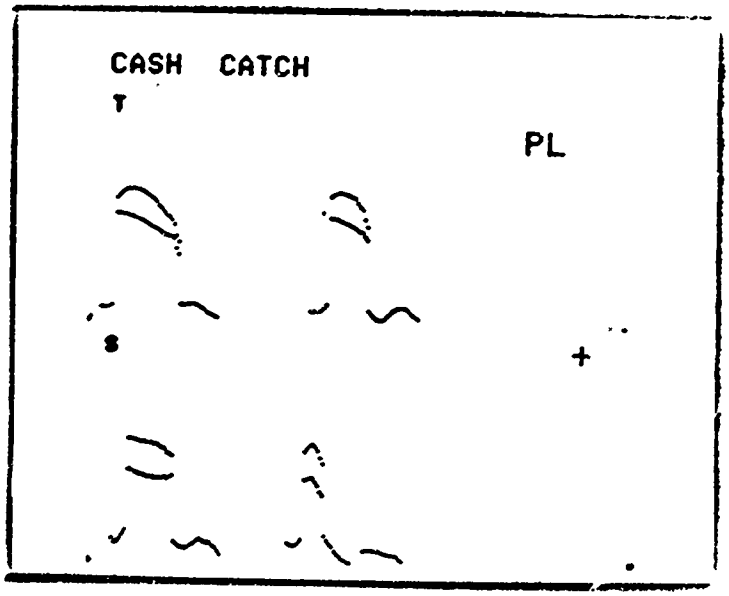


Figure 20b

The last section of this word list has to do with the distinction between the consonant SH and the consonant TCH. Here, the distinctions are very easy to spot. Both of these consonants are produced with the vocal cords not vibrating, and, hence, the

distinctions on the display will be seen in the loudness trace alone, without the presence of the pitch trace below it. The distinctions will thus come in the shape and time course of the loudness function itself. In Figure 20A, the teacher's version of the consonant SH is characterized by a fairly smooth decline in the loudness function after the vocal cords have ceased activity. The consonant TCH is shown by a little bit of loudness after the vocal cords stop,a definite "notch" in the loudness caused by the T in the consonant where the vocal tract is closed completely, and then a final bit of noise that also grades off gradually. the student's attempt to match this has failed because, while the first word seems adequate, the second word does not have the loudness discontinuity required. As Figure 20B shows, it is not necessary that there actually be a hole in the loudness function. All that is required is that the distinction come in the direction of a depression in the loudness function for the consonant TCH. In Figure 20B, the student's match is a little bit better because at least his second word has a larger discontinuity than does his first. It probably would have been better still if the first word had had a gradual, but continuous, decline rather than the slight loudness depression that is seen.

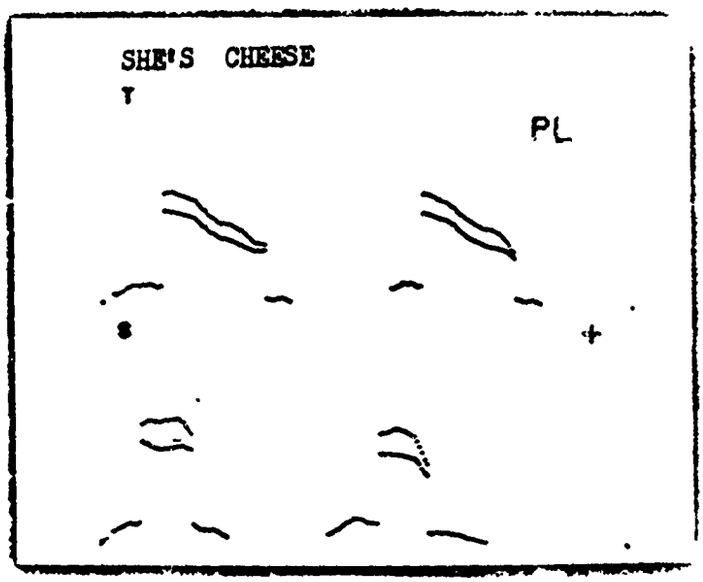


Figure 21

When this consonant distinction occurs at the beginning of the word, it is sometimes difficult to detect the difference between the two loudness traces. The major difference in the teacher traces for these two consonants is that the consonant SH produces a longer period of loudness before the onset of the voice than does the consonant TCH. In addition, one will often find that the consonant TCH produces a larger volume of loudness than the amount of consonant SH at the beginning of the trace. The subject, therefore, has not produced an adequate match, since the duration of the loudness trace before the onset of voicing is in the opposite direction from that of the teacher. Also, the total duration of the voiced section of his words is too small in comparison to that of the teacher.

Here is a suggestion that you might use in attempting to improve your pronunciation of this type of word pair, where the consonant distinction appears at the beginning of the word. Try to add the word "A" before each of the members of the word pair, to allow your voice to start before the word in question. Then, the consonant SH will start immediately after the word A, and the loudness should have a smaller notch in it than it will if the consonant CH immediately follows the word A. This does not make too much sense in terms of English, and, hence, it was not put in the word list itself; but it may help you in producing the distinction a little bit more obviously.

The first part of the third consonant word list has to do with the distinction between the consonants "S" and "Z." To produce both of these consonants, the organs of speech are placed in about the same positions, that is, the tongue is raised immediately behind the front top teeth, and, as air rushes through the small space left, it makes some noise. This is all that is required to produce the "S" sound. Things remain basically the same for the "Z," except

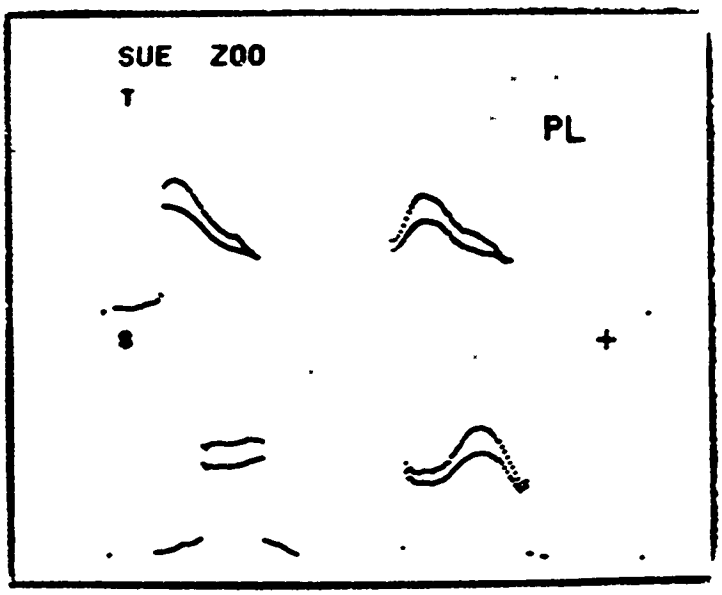


Figure 22

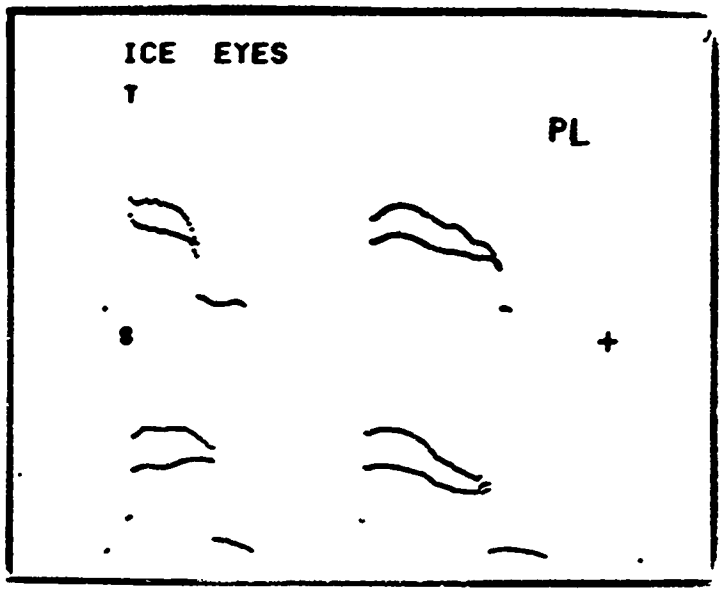


Figure 23

that the vocal cords also vibrate. With the display you have been working with, the distinction caused by the presence or absence of vocal-cord activity is as simple as a one- versus a two-line display for that point in time. When the S sound is produced, the only function to appear is loudness, and that will be shown directly above the invisible baseline connecting the two extreme dots. When the "Z" is produced, the noise part of it will produce a loudness function, but that will appear above the line whose height is determined by the voice pitch used in the production of the "Z" sound. In Figure 22, the student has made the right distinction between two initial consonants. Note that his first word starts out with loudness alone, and then the pitch starts. His second word starts with both pitch and loudness present, indicating that

the Z was spoken. There are some small differences in the shape and the timing of the student's utterances as compared with the teacher's, and these might well be worked on in the future, but the basic distinction appears to be correct from the display. In Figure 23, the student has been less successful, though his utterance is by no means completely incorrect. Here, the consonant distinction occurs at the end of the word. Therefore, words ending with the "S" sound should have a fairly short section of voiced display, followed by a fairly long section of loudness alone, showing the noise produced in the absence of vocal-cord activity. Words ending in "Z" should have a relatively longer period of voicing and a relatively shorter period of loudness-alone at their conclusions. (There may be words in which this loudness-alone section is not present at all.) The student's first word appears to be correct. His second word, while it does have a longer period of voicing, appears to have just as much loudness-alone at its conclusion as does his first word, which ended in the S. This will have to be eliminated before the match appears to be completely correct.

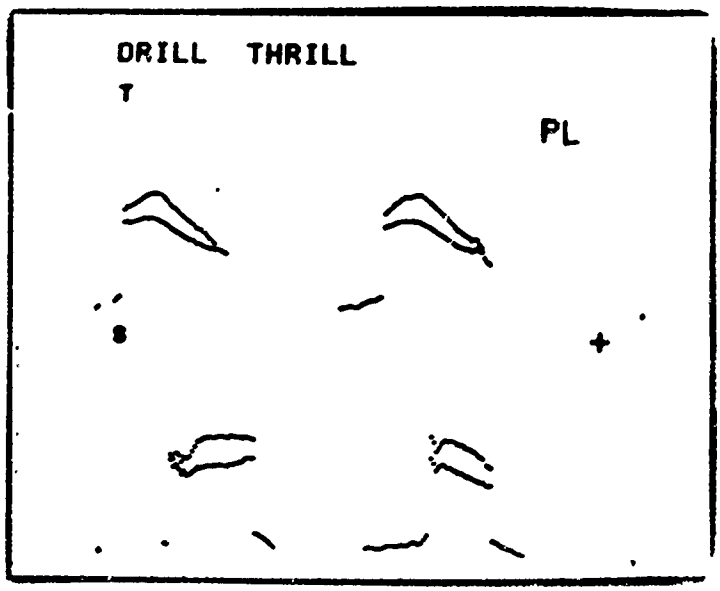


Figure 24

The next consonant distinction treated in the third word list is fairly straightforward. As shown in Figure 24, when consonants to be distinguished appear at the beginning of the words, the presence or absence of voicing at the beginning, or near the beginning, of the word produces a large difference between the displays at the start of each of the words. The teacher's "D" consonant produces only a small amount of loudness before the vocal cords begin activity; the teacher's "TH" produces a much larger amount of voiceless loudness before the rest of the word commences. The student has performed an adequate match despite the fact that his "D" is slightly different than that of the teacher. Note the absence of a large amount of loudness before the pre-voiced D of the student's first word. Recall that it is acceptable in English to pre-voice consonants like D, and refer to discussion above for a description of what a pre-voiced consonant looks like. You will see that the student's D in this case is pre-voiced and acceptable. The student's version of the TH initial consonant is characterized by a large amount of voiceless loudness before the voiced section of the display, and, hence, since the amount of unvoiced loudness is greater for the second than for the first word and roughly comparable to that of the teacher's in extent, the match is adequate. The pitch contour leaves a bit to be desired, and the presence of unvoiced loudness at the conclusion of the words is slightly odd, but the basis distinction at the beginning of the word has been produced, and therefore the match is acceptable.

When the consonant distinction occurs at the end of the word, as in the above two examples, it is fairly easy to detect from the display. First, the vowels preceding the consonant D are lengthened with respect to the way the same vowel would appear when immediately preceding the consonant TH (voiceless). Second, the mouth closure produced by the D at the conclusion of the vowel will make for a

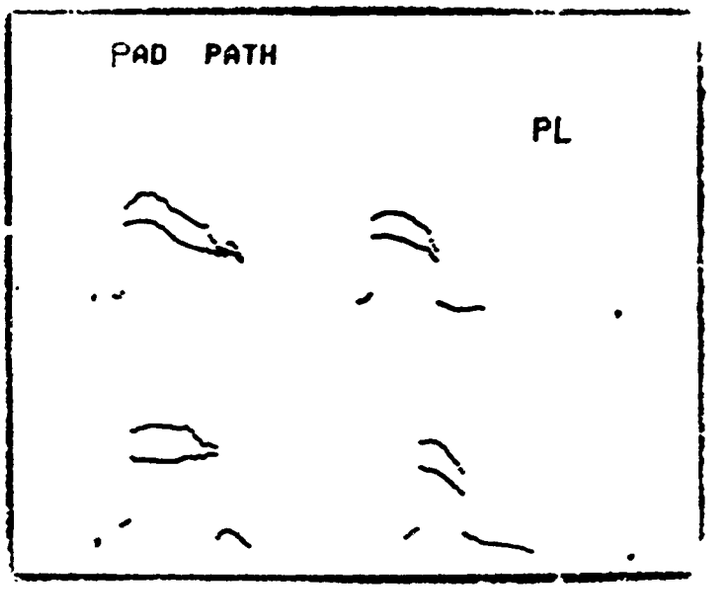


Figure 25

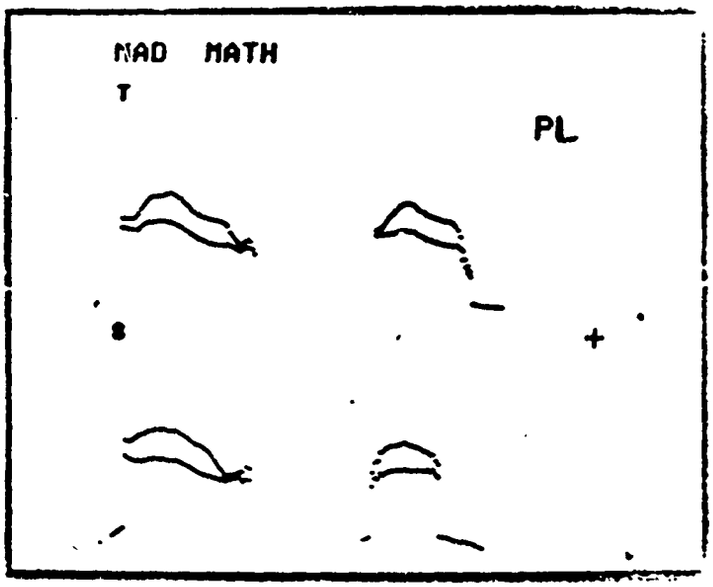


Figure 26

lessening of the loudness function during the presence of voicing, as shown by the words "pad" and "mad," spoken by the teacher in the two above examples. The voiceless TH appears at the conclusion of the words as single-loudness trace in the absence of the voicing line. Even though the student's attempt at the word "pad" in Figure 25 seems odd because of the voiceless loudness bump at the conclusion of the word, it is still an acceptable rendition of the consonant D, because of the fact that the loudness function drops sharply at the conclusion of the voiced part of the word. Some breath must have escaped the speaker's lips at the conclusion of the consonants, but is was nevertheless an acceptable version. The TH following his second word also shows a long trail of loudness at the conclusion of the voiced section. The timing and pitch contours are also about right. Figure 26 shows an even better job

on the part of the student. The final D matches the teacher's final D rather closely, and the same thing goes for the final TH on the student's second word.

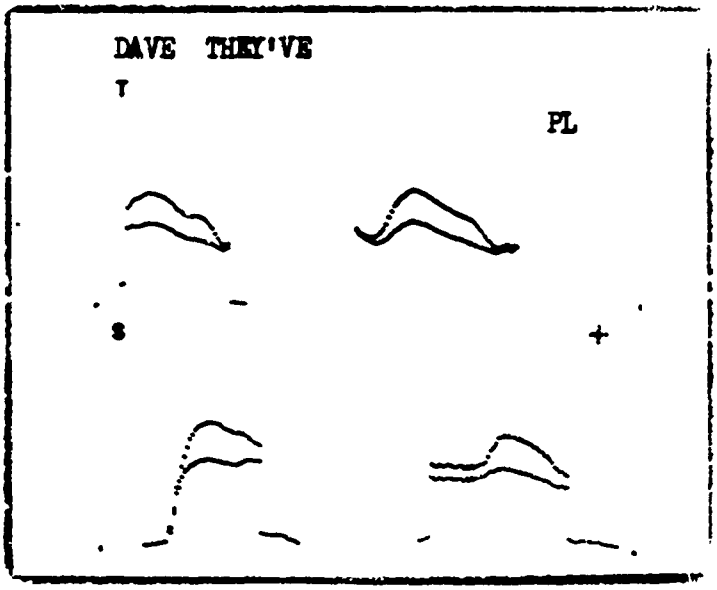


Figure 27

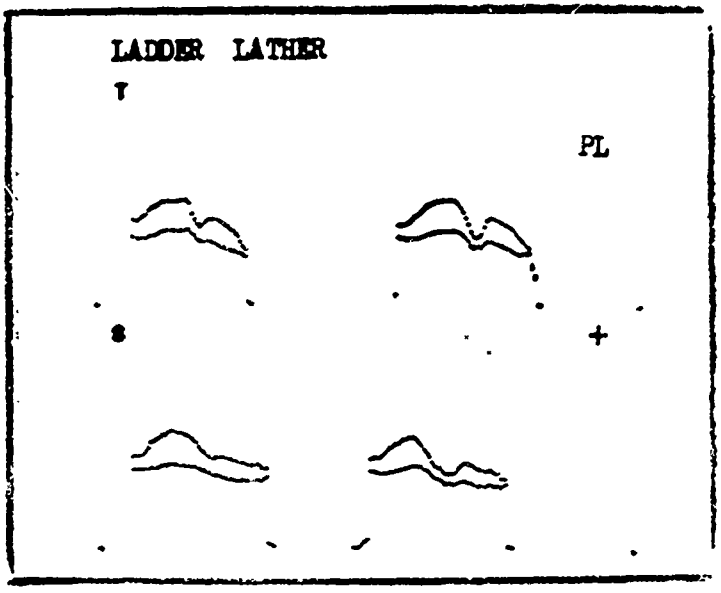


Figure 28

When the consonant D is contrasted with the consonant TH (voiced), the major difference is not the presence or absence of voicing as it was in the above examples; rather, it comes in the fact that the consonant TH (voiced) takes more time than does the consonant D when it occurs at the beginning of the word. When the consonant contrast occurs in the middle or at the end of words, the less drastic vocal-tract closure of the TH (voiced) consonant shows itself in a smaller perturbation of both the pitch and the loudness traces. In Figure 27, the gradual onset of the loudness in the second word of the teacher utterance is a signal for the presence of the TH (voiced) initial consonant, in contrast with the rather fast onset of both traces in the first teacher word.

While the student's second word does have a rather gradual onset, it seems, from the visual display, that the initial consonant from the second word has been drawn out too far. The student's initial consonant distinction has, however, gone in the right direction, because the onset of both pitch and loudness is more gradual in the second word than it is in the first. The match is by no means perfect,though.

When the consonant contrast occurs in the middle of the word, as in the last figure, the major visual distinction between the two of them comes in inspection of the continuity of both traces. In the second teacher word, both traces undergo more gradual changes at the points of interest than they do in the comparable point in time of the first word. The student's attempt at a match has been moderately successful, although it seems that his D in the first word has not been quite as sharply defined as the teacher's. His TH (voiced) production has been slightly drawn out, also. Since the normal error for you will probably be to err in the direction of the TH (voiced), the most important distinction for you to attempt to produce is a sharp discontinuity in both pitch and loudness traces for words having the D consonant.

We hope that these comments will be of some usefulness to you, the student. Please do not hesitate at any time to ask the Monitor for additional assistance or explanations. We hope that you will enjoy and profit from your use of the Automated Pronunciation Instructor system.

APPENDIX 5

STUDENT COMMENTS

ALFREDO BILD

A REPORT ABOUT THE LANGUAGE COMPUTER

I believe that the language computer is a device which is very interesting and useful. More than anything it is a different kind of help. It is not a mechanical teacher, but a place where one can discover, learn, and know the words and languages which are heard everyday.

I don't know why all the people who speak Spanish have the same problem. When we hear the words pronounced, we are not able to understand the sounds. For example; if we hear a word which begins with the letter a; we hear the machine pronounce the ae sound, we say the word as we think we hear it. It seems wrong, but as we reread it we pronounce the word with a then it seems correct.

POINTS IN ITS FAVOR.

I believe that this new invention shouldn't be able to just translate and teach English. It should not only be able to pronounce the English language but should also be able to say the Spanish shounds, because in Spanish there are many sounds also which are not pronounced right.

When we learn to pronounce an s or a z in English, we can use the same pronunciation in Spanish and improve in Spanish. For example: in Spanish the word manzana or zapato would be written with a z but my pronunciation was with an s. When I pronounced sapato, I learned to say the sound z in English, then I use the same in Spanish and say zapato.

I don't believe that I have enough experience to judge the language computer.

APPRENTICESHIP.

I believe that at this point I have not learned anything since I've been here. I have been studying in this university for a month and a half and the English I know now is the same as when I got here. But one day talking to my roommate in English I could understand him where as at first I couldn't, so I knew then that I had learned something.

I believe that unless we dedicate ourselves to studying with the machine, the machine cannot serve its purpose and we have wasted many hours of work in which professionals have spent time sketching, designing to help us with our pronunciation.

ALFREDO BILD (cont)

We are the guinea pigs of this project, the experts understand the functioning of this language computer and correlate the results. They are the only ones to say if the computer is effective or not.

I will feel proud if this project is a success because I was a part of it.

CONCLUSION.

I believe that I'm learning with this machine but only time will tell how effective it really is. Now it is easier for me to prnounce words than before.

TERESA DEMOYA

The system for proving the Spanish pronunciation in the principal language in this world is quite efficient and interesting.

Efficient because the professor and the student work in, are equal level. I mean if the student wants to drill over and over the phrase which is using at the moment he can do it; thus he repeats it as many times as he likes, and the only way that he can learn any language is in the form which I have referred to before.

Interesting because the student participates very much. He has to pay attention, concentrate in the pronunciation trying to produce it; in this way he doesn't get bored.

As I can see the method helps the student much; because due to it, the student can distinguish the sound between two errors which are very similar and thus his ears get accustomed to listen to the words with the proper pronunciation.

The student doesn't know how very important this process is. For this reason he has to try to benefit as much as he can and for getting it he must dedicate time for it to improve the foreign languages.

JENNIE FARBEROFF

In my opinion that machine is a very good help for the Spanish speaking people. In my particular case, I think since I am attending those classes that have the machine, I have improved a lot of my pronunciation, and also my intonation because I have tried to do the same sounds as the machine does.

I can see that I have improved a lot also because now when I hear a word, for example, in the laboratory of English language, I can understand better. Also my pronunciation is better because in the machine I see the spelling of many words and the pronunciation of them; therefore this has helped me a lot in reading and speaking. I am sure that this experiment will succeed because it is obvious that it helps.

JUAN GOMEZ

To me the computer was a creation in order to help the students to better themselves in the pronunciation of the English language.

In cooperation with the teacher, the student can see which are the weak points in his pronunciation, and the teacher can help him with his ears and eyes. In this way he can do better.

ANTONIO GONGORA

I think this method is a very good way to learn English. In this method we learn to pronounce and to distinguish the differences between the consonant and vowel sounds. We also learn soundings, rhyme and tone. Also we look at our own sounds on the screen and listen on the tapes, so we know our own mistakes and we can correct them immediately. This way we know when we speak correctly and we get confidence in ourselves to speak to the people.

I think it would be better to have one hour of this class everyday.

CARLOS HERNANDEZ

The exercise in the computer is very interesting for me because I learn the sounds of the English language. The positive things are: a) we work along with the machine; b) the TV is good; c) the help of Mr. Juan Anguita is very good for us; d) the voice of the teacher in the machine is good.

The negative things are: a) more imitation in the sounds (vowels and consonants); b) the time to practice more.

In general, I like the experience very much and I think it is going to serve me well in learning the English language.

Thank you.

BRIGIDA MORELLI

It is very difficult for me to say something about this modern and unique computer created to teach and improve the pronunciation of the students trying to learn the English language.

I feel that it is difficult for the beginner to remember all the sounds without specific instruction on the position of the tongue and lips. After repeating a sound or word several times I can get a match, but is very difficult for me to do it again in the next class because I can't reproduce the computer sound to practice at home.

VICTOR PERALTA

I think that this method of teaching in the pronunciation of the words is very important and necessary for us, the foreign students and especially for us the Latin-American, that have many difficulties in the pronunciation of the words.

This method consists in a computer. This computer teaches us and helps us to pronounce and differentiate the various sounds of each vowel and consonant. This method consists in a display in which the teacher pronounces first the words and there the student tries to imitate him. In the screen appears a diagram made by the teacher on how to pronounce the words, and the student has to imitate that diagram.

By means of this computer; I'm learning to pronounce correctly little by little.

In conclusion; it's an excellent method for the correct pronunciation of the words.

ILDEFONSA POMPONIO

In my opinion, the computer machine is very useful because it lets you know the correct pronunciation and intonation of the English sounds. The student, who is using this audio-visual method, can improve his knowledge by listening very carefully to the professor's voice, and by trying to imitate the pictures shown on the screen.

However, if the student has the opportunity to listen again to the tape at home, his pronunciation will certainly ameliorate faster.

CESAR VEGAS

The Computer

I have been using the computer a half hour per day. I think what I have learned first is the difference in sounds between a person who speaks English as a native language and me. Before using the computer I can't imagine this difference. I think that I have improved my pronunciation, because when you know the difference and you have the instruments to disappear this difference, it is easy to be made, by practicing.

There are sounds that were too difficult for me and I know this difficulty is not too big. Now I have practiced in the use of the computer; I can get more than before. I think it is good to have a g-od use of the machine, repeated.

I think pronunciation is the most important in English, because without this, you can't speak well. Probably you know the word, but if you don't say it correctly, the meaning may not be clear.

Improving my pronunciation by the computer and learning new words in English, I am going to learn fast.

APPENDIX 6

Script used by accent-rating judges on one student's selected test-day utterances

BBN SECOND LANGUAGE PROJECT EVALUATION

TAPE NO.: 20 SECTION: 1 DATE: JUDGE:

IT'S A MAP	
WHAT'S THAT	
FRUIT BASKET	
NO, IT'S NOT A PEN.	
FRUIT BASKET	
I SAW BILL AND JANE.	
WHAT'S THAT	
I SAW BILL AND JANE.	
IS THAT A DOOR	
WHAT'S THAT	
I SAW BILL AND JANE.	
IT'S A MAP	
IS THAT A DOOR	
IS THAT A DOOR	
FRUIT BASKET	
NO, IT'S NOT A PEN.	
IT'S A MAP	
NO, IT'S NOT A PEN.	

BBN SECOND LANGUAGE PROJECT EVALUATION

TAPE NO.: 20 SECTION: 2 DATE: JUDGE:

SEAL SOIL	
DEEP DIP	--------
LOSS LICE	--------
SAFE SURF	--------
DEEP DIP	--------
LUKE LOOK	--------
PEEP PEP	--------
BAKE BACK	--------
COT CUT	--------
DEEP DIP	--------
LUKE LOOK	--------
COT CUT	--------
PEEP PEP	--------
POT POUT	--------
POT POUT	--------
SEAL SOIL	--------
POT POUT	--------
SAFE SURF	--------
PEEP PEP	--------
BAKE BACK	--------
COT CUT	--------
BAKE BACK	--------
SAFE SURF	--------
LOSS LICE	--------
SEAL SOIL	--------
LOSS LICE	--------
LUKE LOOK	--------

BBN SECOND LANGUAGE PROJECT EVALUATION

TAPE NO.: 20 SECTION: 3 DATE: JUDGE:

GAPE CAPE	
DARE TEAR	
GAPE CAPE	
SHIN CHIN	
BAT VAT	
LACY LAZY	
SHIN CHIN	
BAT VAT	
DARE TEAR	
LACY LAZY	
DO THREW	
DO THREW	
SHIN CHIN	
DO THREW	
DARE TEAR	
LACY LAZY	
BAT VAT	
GAPE CAPE	

INSTRUCTIONS TO JUDGES

Your task today is to evaluate English speech produced by students of English whose native language is Spanish. These students were subjects in an experiment testing a teaching machine whose purpose is to improve the pronunciation of English. Each student read a set of "test words" at various times through the experiment. We wish to find out whether the students' pronunciation of those words changed as time went by. The best way to determine whether there were changes is to ask you to judge them, based on your own experience with English as a native speaker. The methods we will use are time-consuming, but your task is simple.

The recordings of the "test words" have been scrambled and collected on to "judgment tapes," one judgment tape for each student. You will listen to each tape, and will assign a numerical grade to each utterance. You will use the integer numbers 0 through 4, with higher numbers given to more fluent speech. We will give you more details on the grading scale later in the instructions. For now, let us assume that your numbers will range from 0 to 4, bad to good, in accordance with the instructions and with your judgment of each utterance. The judgment tapes contain adequate time for you to consider and respond to each item, before the next one is heard.

You will be given an answer booklet for each different judgment tape, for you to write down your evaluations of the student's speech. This booklet is realy a <u>script</u> that tells you what the student was actually attempting to say. It will help you keep your place. Following each item in the script is a blank space within which you are to write your judgment. This

script is especially helpful when the student's version of the intended speech is garbled. By knowing what the student was trying to say, you can better judge how well he succeeded. Make sure that each line receives a written response from you -- either 0, 1, 2, 3, or 4.

Here is a view of what the judging is like for the entire session. There will be 38 judgment tapes played, in total. There will be rest periods between tapes. Each tape has the same format as the others. The first voice you hear will not be that of the student whose utterances are collected on the tape; it will be an identifier for the tape number. Make sure that it corresponds to the tape number written on the top of the next sheet of the judgment booklet. If it does not, tell the operator, because the script will then not agree with the words you hear. At the start, then, the first page of the judgment booklet corresponds to the first section of the tape.

After you have correctly identified the tape number and assured that your judgment booklet is on the right page, you will hear the student for the first time. The operator will play, at random, from the judgment tape, while you simply listen to the student's voice. This is done so that you can get acquainted with the student's voice before you actually begin judging his speech. The tape will then be brought back to the start for the judgments.

Each tape contains three sections. Each judgment book has three pages, one per section. The first and third section contain 18 items for judgment, and the middle one has 27. Since each student repeated the "test words" three times in the

experiment, it follows that the first and third sections consist of scrambled versions of six items, each heard three times; and that the middle section is composed of three scrambled versions of nine items. The total number of judgments you will make for each judgment tape is therefore 63.

And now: What do the judgments mean? How are you to decide what number to give each utterance? First, remember that your native language is English, and that you will have an instant opinion of each of the items you hear, as to how they compare with your internal standards. Trust that opinion most of all. It is what we hired you for. We wish to guide you only in the translation of that opinion into the judgment numbers, and in directing your attention to certain aspects of the speech. Each item is quite short, which makes your job easier since there are fewer aspects of each item that you need to consider in making your judgment. Also, we are asking you to disregard certain irrelevant aspects of the students' speech, since each of the sections tests the capabilities of the teaching machine in modifying only certain aspects of speech. To be specific: the first section is devoted to intonation contour, rhythm, and stress; the second is concerned with vowels; the last has to do with consonants.

SECTION 1: Phrases and Sentences

In this section, as in the rest, if what the student says sounds like fluent English to you, score it 4. If it is less than OK, break down his performance in the following way. First: Try to disregard the way he pronounces his vowels and consonants. He is reading, and this may produce some confusions on his part. The thing we are most interested in soliciting your opinion about is the manner in which he generates the entire sentence or phrase. Does he use the right "tune?" That is, does his voice "sing" the same kind of melody you would use in speaking what he was attempting? Does he ask questions or make statements using the same upward or downward movements of his voice as you do, as a native English speaker? Furthermore: Does he produce the words with the right rhythm and emphasis? Does he stress important words, does he leave silent intervals between parts of sentences when appropriate? It is difficult to specify just how badly he should do to get a 3, a 2, or a 1; you will simply have to use your best judgment. This may change as you gain experience as a judge; don't let this concern you. Other judges will do the same thing, and they will hear the tapes in a different order than you. If the speech sounds really terrible, give it a 0. Try to distribute your judgments through the range from 0 to 4, if possible; but do not do it if the speech really sounds homogeneous.

SECTION 2: Vowels

This and the third section will be simpler to judge. There are fewer things to keep in mind, and more definite rules for assigning numbers to speech. The items spoken by the students are what is called "minimal pairs" of words, words which differ in only one speech sound. In the case of this section, the

different sounds are vowels. Nine different contrasts are tested. Your job is to indicate how well the student produced each of the vowels, and to disregard (as far as that's possible) any problems he has with the consonants. The amount of time the speaker pauses between words is irrelevant.

As before, if you judge the speech to be fluent, give it a 4. When the vowels sound less than OK, consider the following points. Use the 0 response for really terrible speech; and consider the two vowels as "worth" two points each. So, if the student pronounces one word OK and totally misses the other, give him a 2. If you feel he did OK on one word and ought to get part credit for the other, give him a 3. If he deserves part credit on each word, give him a 2, and so on. Remember that correct pronunciation of a vowel is composed of at least two aspects: the right sound quality and the right duration of the sound. There are some vowel sounds where the quality shifts within the vowel: "diphthongs" like the vowels in "pout" and "lice." Remember to pay attention carefully to the vowels and not to the consonants that surround them. Try to distribute your judgments through the range from 0 to 4, if possible; but do not do it if the speech really sounds homogeneous.

SECTION 3: Consonants

Here, you are to attempt to judge only how well the student produces the consonants. Your judgment scripts have underlining to emphasize this. The same general rules apply here: score 4 if the consonants sound fluent; score zero if the speech is really garbled; and give word and part-credit appropriately.

GENERAL COMMENTS

We realize that we are asking you to do a difficult task. We realize further that your grades may change over time. The purpose of the above standards is to provide you with some sort of absolute yardstick, but invariability is hard to come by in human judgments. We realize this too, and have allowed for it; so just try to do as well and as consistently as you can.

We expect that you will work as carefully and as conscientiously as possible. Much hangs in the balance in this experiment, and so we wish you to consider your judgments as carefully as possible within the time available. Try to provide your full attention to each utterance, disregarding any extraneous sounds that may have remained on the judgment tapes.

There will be speakers whose performance is better than others. Try not to let your scale become relative only to the present speaker, sliding up and down to match the level of each speaker. Try to remain unmoved by swings in ability, but to judge each speaker and indeed each utterance as an independent event. Your increasing experience in this judgment situation may cause some shifts through the entire session; don't become overly concerned with this. If you follow the general guidelines, that is enough for our purposes. Don't try to artificially distinguish between performances that are only slightly different. The categories are fairly broad, and a given level of grading can encompass utterances that differ.

What we are saying is: Try your best to give us a frank impression of how well each speaker produces each utterance -- the better the performance, the higher the score. If you follow the strategy outlined above, we will be satisfied.

BY ALL MEANS ASK ANY QUESTIONS YOU WISH, NOW OR AT ANY TIME DURING THE SESSION.

APPENDIX 8

Comments of Mr. John H. Rogers,
Director of the University of Miami
Intensive English Program, on the API System

July 15, 1974

Dr. Daniel N. Kalikow
Senior Scientist
Bolt Baranek and Newman Inc.
50 Moulton Street
Cambridge, Mass. 02138

Dear Dan:

As you know, I very much regret that it was not possible to keep the API System as part of the instruction offered by Intensive English. Although the result apparently did not show significantly greater improvement made by the subjects than by the control group, I feel that a more intensive use of the API and a personally tailored curriculum might be more effective.

I am certainly impressed with the almost unlimited possibility of the system, and I wish there were some way we could experiment with programming the machine to deal with such problems as L and R confusion with Oriental students, B and P confusion among the Arabs, W and V among the Russian and German students, and the intonational variations we encounter.

Dealing, as we do, with an almost infinite native language population, we have found that pronunciation is a very individual part of the language-learning process. It is hardly practical to try to work with the specific (as opposed to the general) problems and difficulties of individual students on a one-to-one basis. Language laboratories function primarily as pattern practice sections, and pronunciation classes can effectively deal only with the broad spectrum of the phonetic (phonemic) structure of the language. The API could easily be programmed to work, on a one-to-one basis, with those students who demonstrate specific, handicapping problems. A daily period required of selected students, in which minimal pairs or intonational patterns are graphically presented, would be productive at the beginning of the semester. As the student learns to master the intonational pattern or the minimal pairs, different exercises in which the pattern or phoneme is 'buried' within an utterance should be used so that the student learns to recognize and reproduce the material in normal conversatonal situations as well as in fairly structured exercises.

Although I have not seen the results of the experiment, I feel sure that all students improve their pronunciation skill at more or less the same rate while they are beginning the study of the language.

The API would be most effective, for our purposes, at least, with those students who have no 'ear' for the language and could then rely on the visual projection to learn to distinguish meaningful differences and with those students who have developed poor habits either through carelessness or faulty instruction at the beginning of their study.

I am delighted that we have had the opportunity to be part of the experiment, and as I said, I wish it were possible to broaden the base and develop a wider range of material. Should there ever be a possibility in this area, please keep us in mind.

Very truly yours,

John

John H. Rogers, Director
Intensive English Program

JHR/bss

REFERENCES

Balian, J., Language laboratory curriculum materials. Published by the University of Miami. 1972.

Kalikow, D. N. and Stevens, K. N., Technical report No. 1, Second-language learning, 30 June 1971. Submitted to the Advanced Research Projects Agency, Bolt Beranek and Newman, Inc. Report No. 2189.

Kalikow, D. N. and J. A. Swets. Experiments with computer-controlled displays in second-language learning. IEEE Trans. Audio Electroacoust., AU-20, 269-273, 1972.

Kalikow, D. N., Technical report No. 2, Second-language learning, 15 March 1972. Submitted to the Advanced Research Projects Agency. Bolt Beranek and Newman Inc. Report No. 2351.

Kalikow, D. N., and Rollins, A. M., Technical report, Second-language learning, 31 December 1973. Submitted to the Advanced Research Projects Agency. Bolt Beranek and Newman Inc. Report No. 2654.

Lado, R., Test of aural perception in English for Latin-American students. English Language Institute, University of Michigan. 1957.

Nickerson, R. S., and Stevens, K. N., Teaching speech to the deaf: Can a computer help? IEEE Trans. Audio Electroacoust., AU-21, 445-455, 1973.

Wert, J. E., Neidt, C. O., and Ahmann, J. S., Statistical methods in educational and psychological research. New York: Appleton-Century-Crofts, Inc. 1954. viii + 435 pp.